

(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
2018 年 10 月 18 日 (18.10.2018)



(10) 国际公布号
WO 2018/188352 A1

- (51) 国际专利分类号:
G06F 9/50 (2006.01)
- (21) 国际申请号: PCT/CN2017/113773
- (22) 国际申请日: 2017 年 11 月 30 日 (30.11.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710240516.7 2017 年 4 月 13 日 (13.04.2017) 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: 黄平 (HUANG, Ping); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。李升起 (LI, Shengqi); 中国北京市海淀区上地十街 10 号百度大厦三层, Beijing 100085 (CN)。

- (74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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)。

(54) Title: METHOD, DEVICE, AND APPARATUS FOR DETERMINING RESOURCE BALANCE, AND STORAGE MEDIUM

(54) 发明名称: 资源平衡性的确定方法、装置、设备以及存储介质

在设定时间区间内, 统计资源库针对各个资源站点的资源收录量, 以及资源库针对各个资源站点的资源分发量

110

应用目标经济学参数的计算方法, 使用资源收录量以及资源分发量, 计算资源库的资源平衡性参数

120

图 1

- 110 In set time period, compile statistics on volume of resource collected by resource library with respect to resource sites and volume of resource distributed by resource library with respect to resource sites
- 120 By applying calculation method for target economic parameter and using volume of resources collected and volume of resources distributed, calculate resource balance parameter of resource library

(57) Abstract: Disclosed are a method, device, and apparatus for determining resource balance, and a storage medium. The method comprises: in a set time period, compiling statistics on the volume of resources collected by a resource library with respect to resource sites and the volume of resources distributed by the resource library with respect to the resource sites; by applying a calculation method for a target economic parameter and using the volume of the resources collected and the volume of the resources distributed, calculating a resource balance parameter of the resource library. The solution of embodiments of the present invention inventively provides a method for effectively measuring the balance between the collection and distribution of the resources in the resource library, a user is allowed to quantitatively learn, on the basis of the resource balance parameter, whether the collection and distribution of the resources in the resource library are balanced and to adaptively adjust a resource capturing strategy with respect to the resource library on the basis of the calculation result for the resource balance parameter.

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：本发明实施例公开了一种资源平衡性的确定方法、装置、设备及存储介质。所述方法包括：在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量；应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数。本发明实施例的方案创造性的给出了一种有效衡量资源库中资源的收录与分发平衡性的方法，使用户可以根据资源平衡性参数，量化的感知出资源库中资源的收录与分发是否平衡，并可以根据该资源平衡性参数的计算结果，适应性的调整针对所述资源库的资源抓取策略。

资源平衡性的确定方法、装置、设备以及存储介质

本专利申请要求于 2017 年 04 月 13 日提交的，申请号为 201710240516.7，申请人为百度在线网络技术（北京）有限公司，发明名称为“资源平衡性的确定方法、装置、设备以及存储介质”的中国专利申请的优先权，该申请的全文以引用的方式并入本申请中。

技术领域

本发明实施例涉及信息处理技术领域，尤其涉及一种资源平衡性的确定方法、装置、设备以及存储介质。

背景技术

搜索引擎（Search Engine）是指根据一定的策略、运用特定的计算机程序从互联网上搜集信息，在对信息进行组织和处理后，为用户提供检索服务，并将用户检索相关的信息展示给用户的系统，其已经成为广大用户获取网络信息的一项不可或缺的方式。

搜索资源是搜索引擎类产品的基石，一条资源（典型的，一个资源站点中更新的一个网页）从产生到展现给搜索用户要经历资源抓取、入库（也即将资源收录于资源库中）、召回（也即资源的分发）、排序、展现等一系列过程。其中资源的抓取、入库是召回的基础；请求召回的资源数量的多少是对资源抓取、入库质量优良的有效指标，也是影响用户体验的主要因素。在现有技术中，没有对资源库中资源的收录与分发的平衡性进行衡量的方法。

发明内容

本发明实施例提供了一种资源平衡性的确定方法、装置、设备以及存储介质，以有效衡量资源库中资源的收录与分发的平衡性。

第一方面，本发明实施例提供了一种资源平衡性的确定方法，包括：

在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量；

应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数；

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

第二方面，本发明实施例还提供了一种资源平衡性的确定装置，包括：

资源站点信息统计模块，用于在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量；

资源平衡性参数计算模块，用于应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数；

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

第三方面，本发明实施例还提供了一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序所述处理器执行所述程序时实现如本发明实施例中任一所述的资源平衡性的确定方法。

第四方面，本发明实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本发明实施例中任一所述的资源平衡性的确定方法。

本发明实施例提供了一种资源平衡性的确定方法、装置、设备以及存储介质，通过应用衡量经济分配的公平程度的目标经济学参数的计算方法，使用设定时间区间内资源库针对各个资源站点的资源收录量以及资源分发量，计算用于衡量所述资源库的资源收录及分发的平衡性的资源平衡性参数的技术手段，创造性的给出了一种有效衡量资源库中资源的收录与分发平衡性的新方法，使得用户可以根据计算得到的资源平衡性参数，量化的感知出资源库中资源的收录与分发是否平衡，并进而可以根据该资源平衡性参数的计算结果，适应性的调整针对所述资源库的资源抓取策略。

附图说明

图 1 是本发明实施例一中的一种资源平衡性的确定方法的流程图；

图 2 是本发明实施例二中的一种资源平衡性的确定方法的流程图；

图 3a 是本发明实施例三中的一种资源平衡性的确定方法的流程图；

图 3b 是本发明实施例的一种具体应用场景的示意图；

图 4 是本发明实施例四中的一种资源平衡性的确定装置的结构图；

图 5 是本发明实施例五中的一种计算机设备的结构示意图。

具体实施方式

下面结合附图和实施例对本发明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本发明，而非对本发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部结构。

另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部内容。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（或步骤）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。

实施例一

图 1 为本发明实施例一提供的一种资源平衡性的确定方法的流程图，本实施例可适用于对搜索引擎产品资源库的资源收录与资源分发的平衡性进行衡量的情况，该方法可以由本发明实施例提供的资源平衡性的确定装置来执行，该装置可采用软件和/或硬件的方式实现，并一般可集成在搜索引擎产品的服务器中。如图 1 所示，本实施例的方法具体包括：

110、在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量。

在本实施例中，所述资源库具体是指由网络爬虫抓取的，或者由网站自主上报的网页资源构成的，用于进行数据搜索的搜索引擎数据库。其中，所述资源库中包括的资源可以为文本形式的资源、图片形式的资源以及多媒体形式的资源，本实施例对此并不进行限制。

所述设定时间区间可以以设定时间单元（例如，小时、天、一周或者一个月等）为单位进行划分。

所述资源库针对各个资源站点的资源收录量，具体是指所述资源库在该设定时间区间内，对不同资源站点的网页资源的资源收录增量；所述资源库针对各个资源站点的资源分发量，具体是指在该设定时间区间内，搜索引擎产品的召回结果中（针对用户的搜索式返回的搜索结果），与不同资源站点对应的资源数量。

例如：2017.4.10，资源库针对资源站点 1（www.baidu.com）共新收录了 60 个资源，搜索引擎产品共召回了资源库中资源站点 1 的 30 个资源，则在 2017.4.10 这一天，该资源库针对该资源站点 1 的资源收录量为 60，该资源库针对该资源站点 1 的资源分发量为 30。

120、应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数。

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

在本实施例中，发明人通过创造性的劳动发现：在经济学领域中的经济（或者说收入）的分配与资源库中资源的收录及分发具有一定的相似性。因此，可以将设定经济学参数应用于资源的收录及分发领域用于计算资源平衡性参数。

相应的，在本实施例的一个可选的实施方式中，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之前，还可以包括：

预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关

系，并根据所述映射关系，选取所述目标经济学参数。

可选的，可以将所述资源库的资源收录及分发过程与经济学活动之间建立如下映射关系：

将所述资源库映射为整体社会，将所述资源库中包括的资源站点映射为居民；将所述资源库中的资源收录过程映射为收入，将所述资源库中的资源分发过程映射为收入的分配；根据建立的所述映射关系，选取基尼系数作为所述目标经济学参数。

其中，经济学经典指标“基尼系数”的含义是指，在全部居民收入中，用于进行不平均分配的那部分收入所占的比例。基尼系数最大为“1”，最小等于“0”。前者表示居民之间的收入分配绝对不平均，即100%的收入被一个单位的人全部占有了；而后者则表示居民之间的收入分配绝对平均，即人与人之间收入完全平等，没有任何差异。但这两种情况只是在理论上的绝对化形式，在实际生活中一般不会出现。因此，基尼系数的实际数值只能介于0~1之间，基尼系数越小收入分配越平均，基尼系数越大收入分配越不平均。

宏观上，社会全部居民的收入、分配与资源库中站点资源的抓取入库、分发具有很大程度的相似性：

1. 资源库作为资源的容器，可看做基尼系数中的整体社会；
2. 资源站点作为资源的来源，可看做基尼系数中的居民；
3. 入库资源作为资源库的内容来源，可看做基尼系数中的收入；
4. 资源分发作为资源库的向外出口，可看做基尼系数中的分配；

因此，映射到资源层面的基尼系数就可以用来检验资源库中资源的收录及分发是否平等，基尼系数的范围处于0~1之间，基尼系数越小表示资源的收录与分发越平衡，即针对一个资源站点，收录的资源越多，分发的资源越多。

当然，本领域技术人员可以理解的是，鉴于上述映射关系，还可以选取其他类型的经济学参数计算所述资源库的资源平衡性参数，例如：洛伦兹曲线，库兹涅茨曲线，以及泰尔系数等参数，本实施例对此并不进行限制。

本发明实施例提供了一种资源平衡性的确定方法，通过使用用于衡量经济分配的公平程度的目标经济学参数的计算方法，使用设定时间区间内资源库针

对各个资源站点的资源收录量以及资源分发量，计算用于衡量所述资源库的资源收录及分发的平衡性的资源平衡性参数的技术手段，创造性的给出了一种有效衡量资源库中资源的收录与分发平衡性的新方法，使得用户可以根据计算得到的资源平衡性参数，量化的感知出资源库中资源的收录与分发是否平衡，并进而可以根据该资源平衡性参数的计算结果，适应性的调整针对所述资源库的资源抓取策略。

实施例二

图 2 为本发明实施例二提供的一种资源平衡性的确定方法的流程图，本实施例以上述实施例为基础进行具体化，在本实施例中，将应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数，具体为：将所述资源库中包括的资源站点按照资源收录量从小到大的顺序进行排序；按照排序结果，将与各组资源站点分别对应的分发收录比重进行累加求和，得到第一中间量；按照排序结果，将与各组资源站点分别对应的累积分发收录比重进行累加求和，得到第二中间量；根据所述第一中间量以及所述第二中间量，计算所述资源平衡性参数。

相应的，本发明实施例的方法包括：

210、在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量。

220、将所述资源库映射为整体社会，将所述资源库中包括的资源站点映射为居民；将所述资源库中的资源收录过程映射为收入，将所述资源库中的资源分发过程映射为收入的分配。

230、根据所述映射关系，选取基尼系数作为所述目标经济学参数。

其中，所述基尼参数在经济学领域中，具体用于衡量收入分配的公平程度。

240、将所述资源库中包括的资源站点按照资源收录量从小到大的顺序进行排序。

典型的，可以将所述资源库中包括的 N 个资源站点按照资源收录量从小到大的顺序进行排序，并将一个资源站点作为一个独立组，共分为 N 组。

其中，N值与所述资源库中包括的资源站点的总数量相一致。

250、按照排序结果，将与各组资源站点分别对应的分发收录比重进行累加求和，得到第一中间量。

在本实施例中，由于将一个资源站点作为一个独立组，相应的，所述各组资源站点即为各个资源站点。

其中，所述分发收录比重为一个目标资源站点的资源分发量数占全部资源站点的资源分发总量的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积。

典型的，可以根据公式： $A1 = \sum_{i=1}^N Y_i P_i$ ，计算得到第一中间量A1；

其中， Y_i 为第*i*个资源站点的资源分发量数占全部资源站点的资源分发总量的比重； P_i 为第*i*个资源站点的资源收录数占全部资源站点的资源收录总数的比重。

260、按照排序结果，将与各组资源站点分别对应的累积分发收录比重进行累加求和，得到第二中间量。

其中，所述累积分发收录比重为累计到一个目标资源站点的资源收录总数占全部资源站点资源收录总数的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积。

典型的，可以根据公式： $A2 = \sum_{i=1}^N (\sum_{j=1}^i P_j) Y_i$ ，计算得到第二中间量A2。

其中， $(\sum_{j=1}^i P_j)$ 为累计到第*i*个资源站点的资源收录总数占全部资源站点资源收录总数的比重。

270、根据所述第一中间量以及所述第二中间量，计算所述资源平衡性参数。

典型的，可以根据公式： $G = 1 + \sum_{i=1}^N Y_i P_i - 2 \sum_{i=1}^N (\sum_{j=1}^i P_j) Y_i$ ，计算所述资源库的资源平衡性参数G。

本发明实施例仅针对一种基尼系数的计算方法给出了相应的资源平衡性参数的计算方法，当然，本领域技术人员可以理解的是，还可以根据基尼系数的其他计算公式，适应性的调整该资源平衡性参数的计算方法，本实施例对此并不进行限制。

本实施例的技术方案通过将所述资源库中包括的资源站点按照资源收录量

从小到大的顺序进行排序；根据预先建立的资源库的资源收录及分发过程与经济学活动之间的映射关系，对基尼系数的计算公式进行一定的变型处理后，基于该变型后的公式，计算所述资源库的资源平衡性参数，创造性的给出了一种有效衡量资源库中资源的收录与分发平衡性的新方法，使得用户可以根据计算得到的资源平衡性参数，量化的感知出资源库中资源的收录与分发是否平衡，并进而可以根据该资源平衡性参数的计算结果，适应性的调整针对所述资源库的资源抓取策略。

实施例三

图 3a 为本发明实施例三提供的一种资源平衡性的确定方法的流程图，本实施例以上述实施例为基础进行具体化。在本实施例中，将在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量，具体为：在设定时间区间内，监控所述资源库，获取所述资源库中各个资源站点的资源收录增量；以资源站点为单位，对与各资源站点的资源收录增量对应的资源内容进行去重处理；将去重后的结果作为所述资源库针对各个资源站点的资源收录量；在所述设定时间区间内，解析线上日志，获取所述资源库针对各个资源站点的资源分发量；

相应的，在应用所述目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之后，还具体包括：根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略。

相应的，本实施例的方法包括：

310、在设定时间区间内，监控所述资源库，获取所述资源库中各个资源站点的资源收录增量。

320、以资源站点为单位，对与各资源站点的资源收录增量对应的资源内容进行去重处理。

在本实施例中，考虑到最终获取的针对各个资源站点的资源收录增量中，可能由于各种原因，包括了与同一资源站点对应的，重复收录的资源，例如，通过网络爬虫抓取了资源站点 1 中的资源 1，又获取了资源站点 1 自主上报的资

源 1，因此，在统计资源收录增量时，重复记录了该资源 1 两次，因此需要首先对该资源站点 1 的资源收录增量对应的资源内容进行去重处理，这样才能保证后续计算时使用的资源收录量的准确性。

330、将去重后的结果作为所述资源库针对各个资源站点的资源收录量。

340、在所述设定时间区间内，解析线上日志，获取所述资源库针对各个资源站点的资源分发量。

其中，所述线上日志，具体可以是与搜索引擎产品基于用户输入的搜索式召回的搜索结果对应的日志。通过解析记录有所述搜索结果的线上日志，可以进而统计所述设定时间区间内，各个资源站点的资源分发量。

350、应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数。

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

360、根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略。

在本实施例中，在计算得到所述资源平衡性参数之后，可以根据对应的目标经济学参数的经济学含义，对应调整所述资源库的资源抓取策略。

在本实施例的一个可选的实施方式中，根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略可以包括：

如果所述资源平衡性参数的计算结果不满足平衡阈值条件，则获取资源收录量与资源分发量之间的差值超过设定门限的异常资源站点；

根据所述异常资源站点的资源收录量与资源分发量之间的差异类型，对所述异常资源站点的资源抓取策略进行适应性调整。

例如，可以设定一个平衡阈值条件为：0.4-0.6，如果计算得到的资源平衡性参数满足该平衡阈值条件，则可以确定当前的资源抓取策略比较合理，资源的收录及分发过程比较平衡；如果计算得到的资源平衡性参数不满足该平衡阈值条件，则可以确定当前的资源抓取策略不太合理，进而可以获取资源收录量与资源分发量之间的差值超过设定门限（例如，资源收录量-资源分发量大于

1000, 或者资源分发量-资源收录量大于 1000 等) 的异常资源站点。

相应的, 根据所述异常资源站点的资源收录量与资源分发量之间的差异类型 (资源收录量大于资源分发量, 或者资源分发量大于资源收录量), 对所述异常资源站点的资源抓取策略进行适应性调整 (例如: 增大或者减小对所述异常资源站点的抓取频率, 和/或抓取深度等)。

本发明实施例的技术方案通过在设定时间区间内, 监控所述资源库, 获取所述资源库中各个资源站点的资源收录增量; 以资源站点为单位, 对与各资源站点的资源收录增量对应的资源内容进行去重处理; 将去重后的结果作为所述资源库针对各个资源站点的资源收录量; 在所述设定时间区间内, 解析线上日志, 获取所述资源库针对各个资源站点的资源分发量的技术手段, 可以进一步保证资源收录量与资源分发量的计算结果的准确性, 进而提高了后续资源平衡性参数的计算准确性; 同时, 通过根据所述资源平衡性参数的计算结果, 调整针对所述资源库的资源抓取策略的技术手段, 可以达到逐步对资源库的资源抓取策略进行优化的技术效果, 以使所述资源库的资源收录以及分发过程日趋平衡。

具体的, 在图 3b 是本发明实施例的一种具体应用场景的示意图。如 3b 所示, 所述方法具体包括下述流程:

- 1) 监控资源库, 获取天级站点 (按天进行更新的站点) 建库资源量 (也即: 资源收录量) 变化详情;
- 2) 解析线上日志, 获取各站点资源天级分发详情;
- 3) 将 N 个站点按建库资源数量从低到高排序, 一个站点作为独立一组, 分为 N 组;
- 4) 采用基尼系数计算公式计算所述资源平衡性参数。
- 5) 将 4) 中计算的基尼系数变化趋势与资源抓取入库、分发详情 (天级) 结合起来, 引导资源抓取方向调整。

通过上述具体流程, 使得计算得到的基尼系数趋势对资源抓取方向有一定的指导作用, 在此影响下, 可以逐渐改善资源的建库以及资源的分发等性能。

实施例四

图 4 为本发明实施例四提供的一种资源平衡性的确定装置的结构示意图，如图 4 所示，所述装置包括：资源站点信息统计模块 410 以及资源平衡性参数计算模块 420，其中：

资源站点信息统计模块 410，用于在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量。

资源平衡性参数计算模块 420，用于应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数；

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

本发明实施例提供了一种资源平衡性的确定装置，通过使用用于衡量经济分配的公平程度的目标经济学参数的计算方法，使用设定时间区间内资源库针对各个资源站点的资源收录量以及资源分发量，计算用于衡量所述资源库的资源收录及分发的平衡性的资源平衡性参数的技术手段，创造性的给出了一种有效衡量资源库中资源的收录与分发平衡性的新方法，使得用户可以根据计算得到的资源平衡性参数，量化的感知出资源库中资源的收录与分发是否平衡，并进而可以根据该资源平衡性参数的计算结果，适应性的调整针对所述资源库的资源抓取策略。

在上述各实施例的基础上，还可以包括，映射及参数选取模块，用于：

在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之前，预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数。

在上述各实施例的基础上，所述映射及参数选取模块，具体可以用于：

将所述资源库的资源收录及分发过程与经济学活动之间建立如下映射关系：

将所述资源库映射为整体社会，将所述资源库中包括的资源站点映射为居民；将所述资源库中的资源收录过程映射为收入，将所述资源库中的资源分发

过程映射为收入的分配；

根据建立的所述映射关系，选取基尼系数作为所述目标经济学参数。

在上述各实施例的基础上，所述资源平衡性参数计算模块，具体可以用于：
将所述资源库中包括的资源站点按照资源收录量从小到大的顺序进行排序；

按照排序结果，将与各组资源站点分别对应的分发收录比重进行累加求和，
得到第一中间量；

其中，所述分发收录比重为一个目标资源站点的资源分发量数占全部资源站点的资源分发总量的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

按照排序结果，将与各组资源站点分别对应的累积分发收录比重进行累加求和，得到第二中间量；

其中，所述累积分发收录比重为累计到一个目标资源站点的资源收录总数占全部资源站点资源收录总数的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

根据所述第一中间量以及所述第二中间量，计算所述资源平衡性参数。

在上述各实施例的基础上，所述资源站点信息统计模块，具体可以用于：

在设定时间区间内，监控所述资源库，获取所述资源库中各个资源站点的资源收录增量；

以资源站点为单位，对与各资源站点的资源收录增量对应的资源内容进行去重处理；

将去重后的结果作为所述资源库针对各个资源站点的资源收录量；

在所述设定时间区间内，解析线上日志，获取所述资源库针对各个资源站点的资源分发量。

在上述各实施例的基础上，还可以包括，抓取策略调整模块，用于：

在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之后，根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略。

在上述各实施例的基础上，所述抓取策略调整模块，具体可以用于：

如果所述资源平衡性参数的计算结果不满足平衡阈值条件，则获取资源收录量与资源分发量之间的差值超过设定门限的异常资源站点；

根据所述异常资源站点的资源收录量与资源分发量之间的差异类型，对所述异常资源站点的资源抓取策略进行适应性调整。

上述资源平衡性的确定装置可执行本发明任意实施例所提供的资源平衡性的确定方法，具备执行的资源平衡性的确定方法相应的功能模块和有益效果。

实施例五

图 5 为本发明实施例五提供的一种计算机设备的结构示意图。图 5 示出了适于用来实现本发明实施方式的示例性计算机设备 12 的框图。图 5 显示的计算机设备 12 仅仅是一个示例，不对本发明实施例的功能和使用范围带来任何限制。

如图 5 所示，计算机设备 12 以通用计算设备的形式表现。计算机设备 12 的组件可以包括但不限于：一个或者多个处理器或者处理单元 16，系统存储器 28，连接不同系统组件（包括系统存储器 28 和处理单元 16）的总线 18。

总线 18 表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构（ISA）总线，微通道体系结构（MAC）总线，增强型 ISA 总线、视频电子标准协会（VESA）局域总线以及外围组件互连（PCI）总线。

计算机设备 12 典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备 12 访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质。

系统存储器 28 可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器（RAM）30 和/或高速缓存存储器 32。计算机设备 12 可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储系统 34 可以用于读写不可移动的、非易失性磁介质（图 5 未显示，

通常称为“硬盘驱动器”)。尽管图 5 中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如 CD-ROM, DVD-ROM 或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线 18 相连。存储器 28 可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本发明各实施例的功能。

具有一组(至少一个)程序模块 42 的程序/实用工具 40,可以存储在例如存储器 28 中,这样的程序模块 42 包括——但不限于——操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块 42 通常执行本发明所描述的实施例中的功能和/或方法。

计算机设备 12 也可以与一个或多个外部设备 14(例如键盘、指向设备、显示器 24 等)通信,还可与一个或者多个使得用户能与该计算机设备 12 交互的设备通信,和/或与使得该计算机设备 12 能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口 22 进行。并且,计算机设备 12 还可以通过网络适配器 20 与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器 20 通过总线 18 与计算机设备 12 的其它模块通信。应当明白,尽管图 5 中未示出,可以结合计算机设备 12 使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储系统等。

处理单元 16 通过运行存储在系统存储器 28 中的程序,从而执行各种功能应用以及数据处理,例如实现本发明实施例所提供的资源平衡性的确定方法。

也即,所述处理单元执行所述程序时实现:在设定时间区间内,统计资源库针对各个资源站点的资源收录量,以及所述资源库针对各个资源站点的资源分发量;应用目标经济学参数的计算方法,使用所述资源收录量以及所述资源分发量,计算所述资源库的资源平衡性参数;其中,所述目标经济学参数用于衡量经济分配的公平程度,所述资源平衡性参数用于衡量所述资源库的资源收

录及分发的平衡性。

进一步地，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之前，还可以包括：

预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数。

进一步地，预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数，可以包括：

将所述资源库的资源收录及分发过程与经济学活动之间建立如下映射关系：

将所述资源库映射为整体社会，将所述资源库中包括的资源站点映射为居民；将所述资源库中的资源收录过程映射为收入，将所述资源库中的资源分发过程映射为收入的分配；

根据建立的所述映射关系，选取基尼系数作为所述目标经济学参数。

进一步地，应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源平衡性参数，可以包括：

将所述资源库中包括的资源站点按照资源收录量从小到大的顺序进行排序；

按照排序结果，将与各组资源站点分别对应的分发收录比重进行累加求和，得到第一中间量；

其中，所述分发收录比重为一个目标资源站点的资源分发量数占全部资源站点的资源分发总量的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

按照排序结果，将与各组资源站点分别对应的累积分发收录比重进行累加求和，得到第二中间量；

其中，所述累积分发收录比重为累计到一个目标资源站点的资源收录总数占全部资源站点资源收录总数的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

根据所述第一中间量以及所述第二中间量，计算所述资源平衡性参数。

进一步地，在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量，可以包括：

在设定时间区间内，监控所述资源库，获取所述资源库中各个资源站点的资源收录增量；

以资源站点为单位，对与各资源站点的资源收录增量对应的资源内容进行去重处理；

将去重后的结果作为所述资源库针对各个资源站点的资源收录量；

在所述设定时间区间内，解析线上日志，获取所述资源库针对各个资源站点的资源分发量。

进一步地，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之后，还可以包括：

根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略。

进一步地，根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略，可以包括：

如果所述资源平衡性参数的计算结果不满足平衡阈值条件，则获取资源收录量与资源分发量之间的差值超过设定门限的异常资源站点；

根据所述异常资源站点的资源收录量与资源分发量之间的差异类型，对所述异常资源站点的资源抓取策略进行适应性调整。

实施例六

本发明实施例六提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本发明所有发明实施例提供的资源平衡性的确定方法：

也即，该程序被处理器执行时实现：在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量；应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数；其中，所述目标经济学参数用于衡量经济

分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

该程序被处理器执行时，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之前，还可以实现：

预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数。

该程序被处理器执行时，预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数，还可以实现：

将所述资源库的资源收录及分发过程与经济学活动之间建立如下映射关系：

将所述资源库映射为整体社会，将所述资源库中包括的资源站点映射为居民；将所述资源库中的资源收录过程映射为收入，将所述资源库中的资源分发过程映射为收入的分配；

根据建立的所述映射关系，选取基尼系数作为所述目标经济学参数。

该程序被处理器执行时，应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源平衡性参数，可以实现为：

将所述资源库中包括的资源站点按照资源收录量从小到大的顺序进行排序；

按照排序结果，将与各组资源站点分别对应的分发收录比重进行累加求和，得到第一中间量；

其中，所述分发收录比重为一个目标资源站点的资源分发量数占全部资源站点的资源分发总量的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

按照排序结果，将与各组资源站点分别对应的累积分发收录比重进行累加求和，得到第二中间量；

其中，所述累积分发收录比重为累计到一个目标资源站点的资源收录总数

占全部资源站点资源收录总数的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

根据所述第一中间量以及所述第二中间量，计算所述资源平衡性参数。

该程序被处理器执行时，在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量，可以实现为：

在设定时间区间内，监控所述资源库，获取所述资源库中各个资源站点的资源收录增量；

以资源站点为单位，对与各资源站点的资源收录增量对应的资源内容进行去重处理；

将去重后的结果作为所述资源库针对各个资源站点的资源收录量；

在所述设定时间区间内，解析线上日志，获取所述资源库针对各个资源站点的资源分发量。

该程序被处理器执行时，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之后，还可以实现：

根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略。

该程序被处理器执行时，根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略，可以实现为：

如果所述资源平衡性参数的计算结果不满足平衡阈值条件，则获取资源收录量与资源分发量之间的差值超过设定门限的异常资源站点；

根据所述异常资源站点的资源收录量与资源分发量之间的差异类型，对所述异常资源站点的资源抓取策略进行适应性调整。

可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子（非穷举的

列表) 包括: 具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (EPROM 或闪存)、光纤、便携式紧凑磁盘只读存储器 (CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中, 计算机可读存储介质可以是任何包含或存储程序的有形介质, 该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号, 其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式, 包括——但不限于——电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质, 该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

计算机可读介质上包含的程序代码可以用任何适当的介质传输, 包括——但不限于——无线、电线、光缆、RF 等等, 或者上述的任意合适的组合。

可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码, 所述程序设计语言包括面向对象的程序设计语言——诸如 Java、Smalltalk、C++, 还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中, 远程计算机可以通过任意种类的网络——包括局域网 (LAN) 或广域网 (WAN)——连接到用户计算机, 或者, 可以连接到外部计算机 (例如利用因特网服务提供商来通过因特网连接)。

通过以上关于实施方式的描述, 所属领域的技术人员可以清楚地了解到, 本发明可借助软件及必需的通过硬件来实现, 当然也可以通过硬件实现, 但很多情况下前者是更佳的实施方式, 基于这样的理解, 本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来, 该计算机

软件产品可以存储在计算机可读存储介质中，如磁盘、光盘、只读存储记忆体（Read-Only Memory, ROM）或随机存储记忆体（Random Access Memory, RAM）等，包括若干指令用以使得一台计算机设备（可以是个人计算机、服务器，或者网络设备等等）执行本发明各个实施例所述的方法。

值得注意的是，上述资源平衡性的确定装置的实施例中，所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的具体名称也只是为了便于相互区分，并不用于限定本发明的保护范围。

注意，上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解，本发明不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此，虽然通过以上实施例对本发明进行了较为详细的说明，但是本发明不仅仅限于以上实施例，在不脱离本发明构思的情况下，还可以包括更多其他等效实施例，而本发明的范围由所附的权利要求范围决定。

权 利 要 求 书

1、一种资源平衡性的确定方法，其特征在于，包括：

在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量；

应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数；

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

2、根据权利要求1所述的方法，其特征在于，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之前，还包括：

预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数。

3、根据权利要求2所述的方法，其特征在于，预先建立所述资源库的资源收录及分发过程与经济学活动之间的映射关系，并根据所述映射关系，选取所述目标经济学参数，包括：

将所述资源库的资源收录及分发过程与经济学活动之间建立如下映射关系：

将所述资源库映射为整体社会，将所述资源库中包括的资源站点映射为居民；将所述资源库中的资源收录过程映射为收入，将所述资源库中的资源分发过程映射为收入的分配；

根据建立的所述映射关系，选取基尼系数作为所述目标经济学参数。

4、根据权利要求3所述的方法，其特征在于，应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源平衡性参数，包括：

将所述资源库中包括的资源站点按照资源收录量从小到大的顺序进行排序；

按照排序结果，将与各组资源站点分别对应的分发收录比重进行累加求和，得到第一中间量；

其中，所述分发收录比重为一个目标资源站点的资源分发量数占全部资源站点的资源分发总量的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

按照排序结果，将与各组资源站点分别对应的累积分发收录比重进行累加求和，得到第二中间量；

其中，所述累积分发收录比重为累计到一个目标资源站点的资源收录总数占全部资源站点资源收录总数的比重与所述目标资源站点的资源收录数占全部资源站点的资源收录总数的比重的乘积；

根据所述第一中间量以及所述第二中间量，计算所述资源平衡性参数。

5、根据权利要求 1-4 任一项所述的方法，其特征在于，在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量，包括：

在设定时间区间内，监控所述资源库，获取所述资源库中各个资源站点的资源收录增量；

以资源站点为单位，对与各资源站点的资源收录增量对应的资源内容进行去重处理；

将去重后的结果作为所述资源库针对各个资源站点的资源收录量；

在所述设定时间区间内，解析线上日志，获取所述资源库针对各个资源站点的资源分发量。

6、根据权利要求 1-4 任一项所述的方法，其特征在于，在应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数之后，还包括：

根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略。

7、根据权利要求 6 所述的方法，其特征在于，根据所述资源平衡性参数的计算结果，调整针对所述资源库的资源抓取策略，包括：

如果所述资源平衡性参数的计算结果不满足平衡阈值条件，则获取资源收录量与资源分发量之间的差值超过设定门限的异常资源站点；

根据所述异常资源站点的资源收录量与资源分发量之间的差异类型，对所述异常资源站点的资源抓取策略进行适应性调整。

8、一种资源平衡性的确定装置，其特征在于，包括：

资源站点信息统计模块，用于在设定时间区间内，统计资源库针对各个资源站点的资源收录量，以及所述资源库针对各个资源站点的资源分发量；

资源平衡性参数计算模块，用于应用目标经济学参数的计算方法，使用所述资源收录量以及所述资源分发量，计算所述资源库的资源平衡性参数；

其中，所述目标经济学参数用于衡量经济分配的公平程度，所述资源平衡性参数用于衡量所述资源库的资源收录及分发的平衡性。

9、一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时实现如权利要求 1-7 中任一所述的方法。

10、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求 1-7 中任一所述的方法。

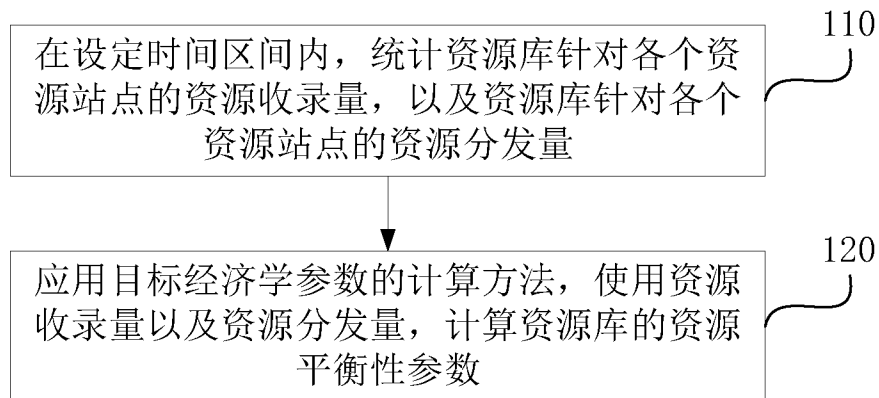
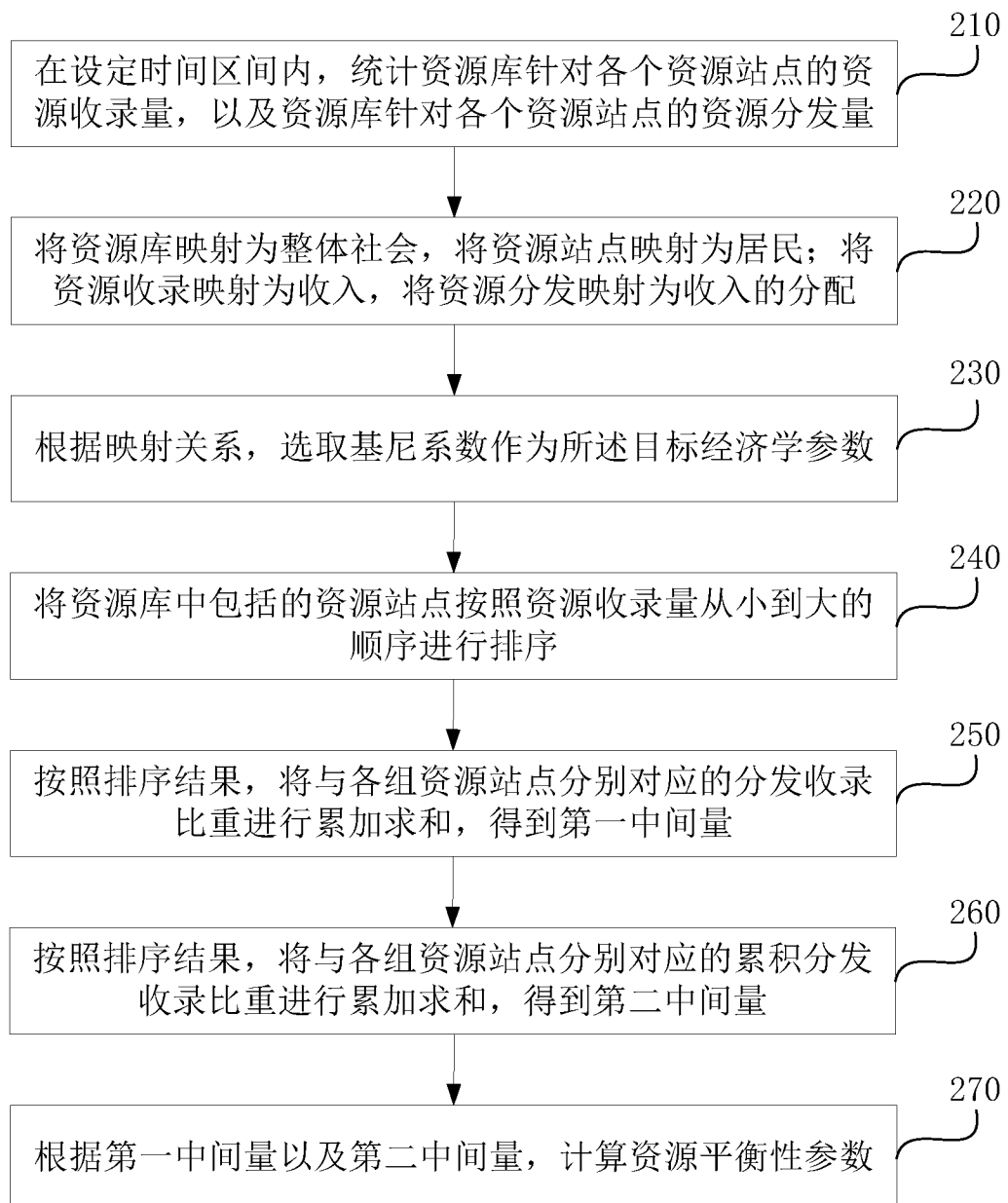


图 1



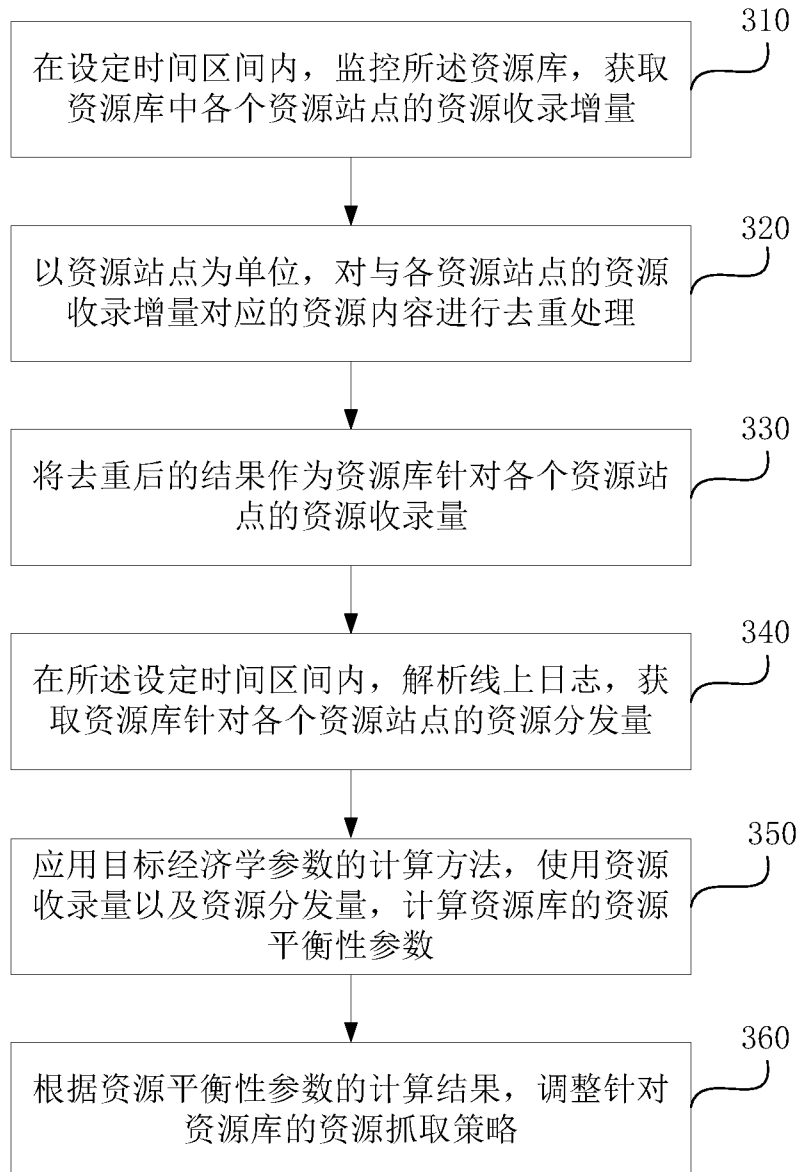


图 3a

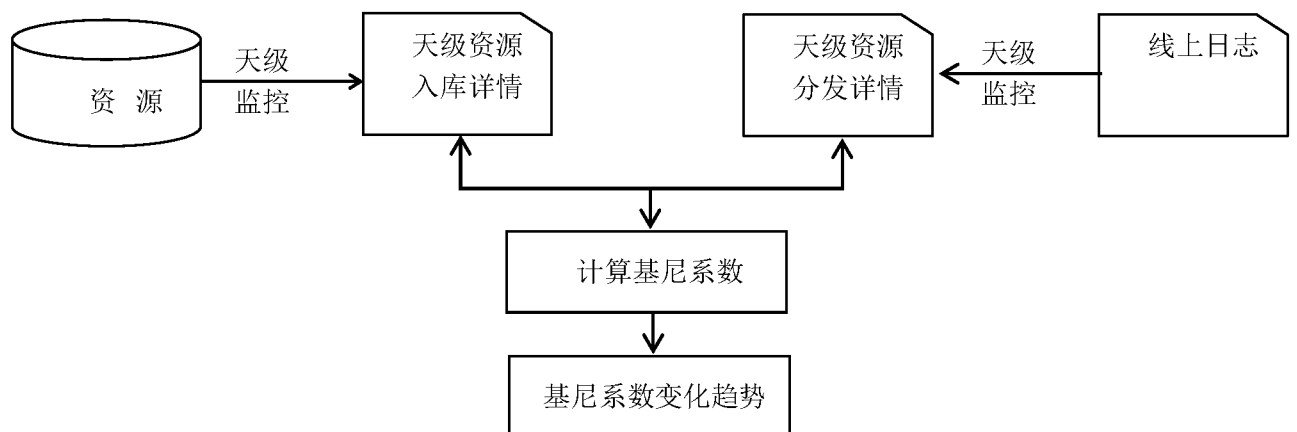


图 3b

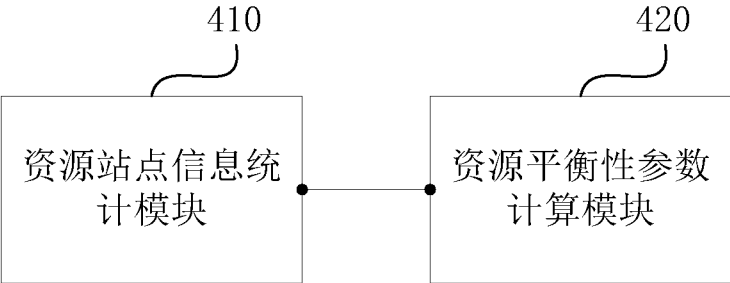


图 4

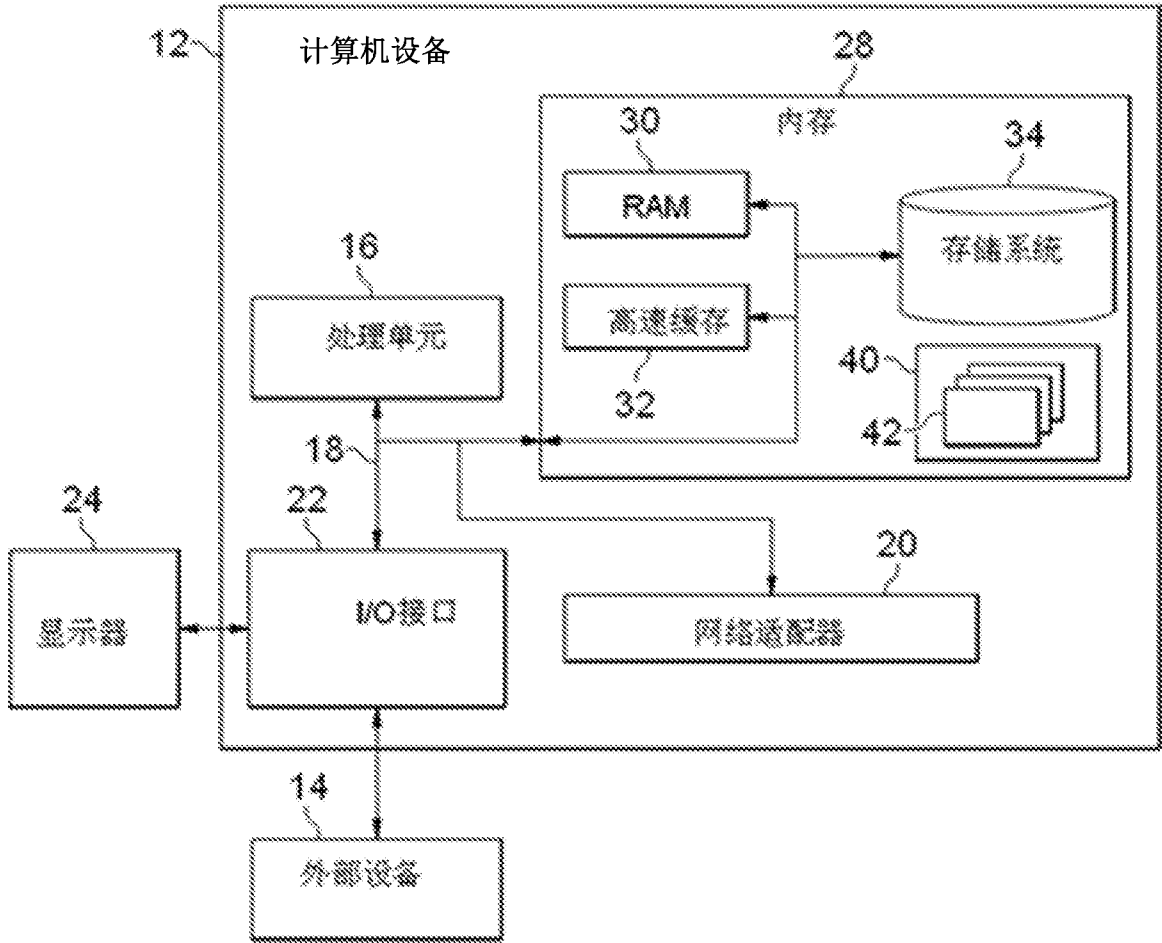


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/113773

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/50 (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)

VEN, CNABS, CNTXT, CNKI: 搜索引擎, 资源, 网页, 页面, 入库, 收录, 分发, 召回, 分配, 调度, 经济学, 基尼, 平衡, 均衡, 公平, 调整, 抓取, 统计, 站点, 网站, search, engine, resource, web page, distribut+, balance, station, +site, count+, quantity, statistic+, economic+, fair+, +equal+, gini

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103888976 A (NORTHEASTERN UNIVERSITY), 25 June 2014 (25.06.2014), entire document	1-10
A	CN 105956247 A (DALIAN UNIVERSITY OF TECHNOLOGY), 21 September 2016 (21.09.2016), entire document	1-10
A	US 2009132336 A1 (YAHOO INC.), 21 May 2009 (21.05.2009), entire document	1-10
A	CN 103646342 A (CUBEAD SEM INC.), 19 March 2014 (19.03.2014), entire document	1-10
PX	CN 107066337 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 18 August 2017 (18.08.2017), claims 1-10	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 03 January 2018	Date of mailing of the international search report 16 January 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer WANG, Yue Telephone No. (86-10) 62089109

International application No.
PCT/CN2017/113773

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/113773

A. 主题的分类

G06F 9/50(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))

VEN, CNABS, CNTXT, CNKI: 搜索引擎, 资源, 网页, 页面, 入库, 收录, 分发, 召回, 分配, 调度, 经济学, 基尼, 平衡, 均衡, 公平, 调整, 抓取, 统计, 站点, 网站, search, engine, resource, web page, distribut+, balan-
ce, station, +site, count+, quantity, statistic+, economic+, fair+, +equal+, gini

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
A	CN 103888976 A (东北大学) 2014年 6月 25日 (2014 - 06 - 25) 全文	1-10
A	CN 105956247 A (大连理工大学) 2016年 9月 21日 (2016 - 09 - 21) 全文	1-10
A	US 2009132336 A1 (YAHOO INC) 2009年 5月 21日 (2009 - 05 - 21) 全文	1-10
A	CN 103646342 A (北京博雅立方科技有限公司) 2014年 3月 19日 (2014 - 03 - 19) 全文	1-10
PX	CN 107066337 A (百度在线网络技术北京有限公司) 2017年 8月 18日 (2017 - 08 - 18) 权利要求1-10	1-10

☐ 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

2018年 1月 3日

国际检索报告邮寄日期

2018年 1月 16日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

王越

电话号码 (86-10)62089109

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/113773

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	103888976	A	2014年 6月 25日	无	
CN	105956247	A	2016年 9月 21日	无	
US	2009132336	A1	2009年 5月 21日	US 7805331 B2	2010年 9月 28日
CN	103646342	A	2014年 3月 19日	无	
CN	107066337	A	2017年 8月 18日	无	

表 PCT/ISA/210 (同族专利附件) (2009年7月)