

(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
2017 年 10 月 5 日 (05.10.2017)



(10) 国际公布号
WO 2017/167063 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/077450
- (22) 国际申请日: 2017 年 3 月 21 日 (21.03.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610195039.2 2016 年 3 月 30 日 (30.03.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Cayman (KY)。
- (72) 发明人: 及
- (71) 申请人 (仅对美国): 刘学 (LIU, Xue) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.);

中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (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, 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)。

[见续页]

(54) Title: DATA RETRIEVAL METHOD AND DEVICE, AND DATA STORAGE METHOD AND DEVICE

(54) 发明名称: 一种数据检索方法和装置、一种数据存储方法和装置

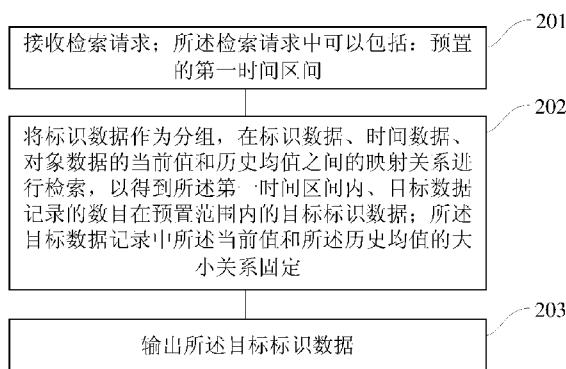


图 2

- 210 Receive a retrieval request, wherein the retrieval request comprises a preset first time interval
- 202 Use identification data for grouping, and perform retrieval on mapping relationships between current values and historical mean values of identification data, time data, and object data, so as to obtain target identification data within the first time interval, wherein the number of target data records are within a preset range and magnitude relationships between the current values and the historical mean values in the target data records are fixed
- 203 Output the target identification data

(57) Abstract: A data retrieval method and device, and a data storage method and device are provided. The retrieval method comprises: receiving a retrieval request, wherein the retrieval request comprises a preset first time interval (201); using identification data for grouping, and performing retrieval on mapping relationships between current values and historical mean values of identification data, time data, and object data, so as to obtain target identification data within the first time interval, wherein the number of target data records are within a preset range, and magnitude relationships between the current values and the historical mean values in the target data records are fixed (202); outputting the target identification data (203). The method and device of the present invention can greatly improve efficiency of trend determination and efficiency of obtaining of target identification data, conserving system resources, and reducing human resource costs.

(57) 摘要: 一种数据检索方法和装置、一种数据存储方法和装置, 其中的检索方法具体包括: 接收检索请求; 所述检索请求中包括: 预置的第一时间区间 (201); 将标识数据作为分组, 在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索, 以得到所述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据; 所述目标数据记录中所述当前值和所述历史均值的大小关系固定 (202); 输出所述目标标识数据 (203)。所述方法和装置能够大大提高趋势判定的效率和目标标识数据的获取效率, 且能够节省系

统资源, 以及能够节省人力成本。

WO 2017/167063 A1

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种数据检索方法和装置、一种数据存储方法和装置

技术领域

本申请涉及信息技术技术领域，特别是涉及一种数据检索方法和装置、一种数据存储方法和装置。

5 背景技术

随着信息技术的高速发展，行业数据量急剧增长。目前各行业涌现的大数据具有数据量大和具有时间相关性的特性，对大数据进行统计分析可以帮助制定行业决策。例如，对于以交易型数据为主的行业，例如金融、电子商务等，对大数据进行分析，能够快速提炼出具有商业价值的信息。其中，常见的一种大数据分析需求具体为：判定对象数据在某个时间段内的增长或者下降趋势。例如，假设网站具有数千万个会员，则需要对所有会员的交易额数据进行分析，以得到在某个时间段内交易额数据呈增长趋势的会员。

一种现有方案可以利用趋势图判定对象数据在某个时间段内的增长或者下降趋势，以借助于趋势图的连续曲线的升降变化，来反映对象数据的动态变化过程。具体地，可以将所有对象数据一次性读出，逐点绘制。

然而，在实际应用中，上述逐点绘制过程需要耗费大量的时长，且占用较多的系统资源。另外，分析人员在查看和分析绘制得到的趋势图后，才能得到对应的趋势判定结果，不仅耗费了大量人力成本，而且影响了趋势判定的效率。

发明内容

本申请实施例所要解决的技术问题是提供一种数据检索方法，能够大大提高趋势判定的效率和目标标识数据的获取效率，且能够节省系统资

源，以及能够节省人力成本。

相应的，本申请实施例还提供了一种数据检索装置、一种数据存储方法和装置，用以保证上述方法的实现及应用。

为了解决上述问题，本申请公开了一种数据检索方法，包括：

5 接收检索请求；所述检索请求中包括：预置的第一时间区间；

 将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据；所述目标数据记录中所述当前值和所述历史均值的大小关系固定；

10 输出所述目标标识数据。

 可选地，所述当前值和所述历史均值的大小关系固定，包括：所述当前值大于所述历史均值，或者，所述当前值小于所述历史均值。

 可选地，所述目标数据记录的数目在预置范围内，包括：所述目标数据记录的数目超过阈值。

15 可选地，所述阈值大于等于所述第一时间区间的长度的二分之一，且所述阈值小于等于所述第一时间区间的长度。

 可选地，所述将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索的步骤，包括：

 将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述对象数据值和所述历史均值的大小关系固定的目标数据记录；

20

 建立并保存对象数据和所述目标数据记录的数目之间的映射关系；

 在所述对象数据和所述目标数据记录的数目之间的映射关系中进行检索，以得到目标数据记录的数目在预置范围内的目标标识数据。

25 可选地，所述历史均值包括：对象数据值在当前时间数据对应的第二

时间区间内的均值，所述第二时间区间包括：早于所述当前时间数据的时间数据。

另一方面，本申请公开了一种数据存储方法，包括：

依据标识数据、时间数据和对象数据的当前值之间的映射关系，确定
5 所述对象数据在当前时间数据对应的第二时间区间内的均值，作为对应的历史均值；所述第二时间区间包括：早于所述当前时间数据的时间数据；

依据所述历史均值，建立并保存标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系。

可选地，所述第二时间区间的终点为前一时间数据。

10 可选地，所述第二时间区间的长度大于等于第一时间区间的长度的二分之一，且所述第二时间区间的长度小于等于第一时间区间的长度。

再一方面，本申请公开了一种数据检索装置，包括：

接收模块，用于接收检索请求；所述检索请求中包括：预置的第一时间区间；

15 检索模块，用于将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据；所述目标数据记录中所述当前值和所述历史均值的大小关系固定；以及

输出模块，用于输出所述目标标识数据。

20 可选地，所述当前值和所述历史均值的大小关系固定，包括：所述当前值大于所述历史均值，或者，所述当前值小于所述历史均值。

可选地，所述目标数据记录的数目在预置范围内，包括：所述目标数据记录的数目超过阈值。

可选地，所述检索模块，包括：

25 第一检索子模块，用于将标识数据作为分组，在标识数据、时间数据、

对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述对象数据值和所述历史均值的大小关系固定的目标数据记录；

保存子模块，用于建立并保存对象数据和所述目标数据记录的数目之间的映射关系；以及

- 5 第二检索子模块，用于在所述对象数据和所述目标数据记录的数目之间的映射关系中进行检索，以得到目标数据记录的数目在预置范围内的目标标识数据。

又一方面，本申请公开了一种数据存储装置，包括：

- 10 确定模块，用于依据标识数据、时间数据和对象数据的当前值之间的映射关系，确定所述对象数据在当前时间数据对应的第二时间区间内的均值，作为对应的历史均值；所述第二时间区间包括：早于所述当前时间数据的时间数据；以及

保存模块，用于依据所述历史均值，建立并保存标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系。

15

与现有方案相比，本申请实施例包括以下优点：

- 20 相对于现有方案中趋势图的绘制过程，由于本申请实施例采用简单的检索逻辑进行符合特定趋势的目标标识数据（例如交易额持续增长的用户标识或网站标识）的检索，故能够大大提高趋势判定的效率和目标标识数据的获取效率，且能够节省系统资源。

另外，相对于现有方案中分析人员在查看和分析绘制得到的趋势图后，才能得到对应的趋势判定结果，本申请实施例能够直接输出符合特定趋势的目标标识数据，因此能够进一步提高趋势判定的效率和目标标识数据的获取效率，且能够节省对应的人力成本。

25

附图说明

图 1 是本申请的一种数据存储方法实施例的步骤流程图；；

图 2 是本申请的一种数据检索方法实施例的步骤流程图；

图 3 是本申请的一种数据检索装置实施例的结构框图；以及

5 图 4 是本申请的一种数据存储装置实施例的结构框图。

具体实施方式

为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

10 针对现有方案中趋势图的绘制耗费大量的时长、占用较多的系统资源、以及趋势图的分析需要耗费大量的人力成本的技术问题，本申请实施例提供了一种依据连续区间内当前值和历史均值的大小关系确定对象数据的趋势的方案，其可以基检索的方式实现，具体地，可以在检索请求中携带预置的第一时间区间，并将标识数据作为分组，在标识数据、时间
15 数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到上述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据，上述目标数据记录中上述当前值和上述历史均值的大小关系固定，进而输出上述目标标识数据。

由于上述历史均值可用于表示对象数据在当前对象数据对应的历史时
20 间区间内，故其能够反映对象数据在上述历史时间区间内的总体情况，故本申请实施例中上述当前值和上述历史均值的大小关系能够反映对象数据当前的增量变化，具体地，若上述当前值大于上述历史均值，则对应的增量为正，若上述当前值小于上述历史均值，则对应的增量为负；因此，上述当前值和上述历史均值的大小关系固定可以反映某种特定趋势（如增
25 长或下降趋势），从而可以依据上述第一时间区间内目标数据记录的数目

确定当前标识数据在第一时间区间的对象数据是否整体上呈现出的特定增长或下降趋势。进一步，在上述第一时间区间内目标数据记录的数目在预置范围内时，可以确定第一时间区间的对象数据整体上呈现出的特定趋势，故可以将当前标识数据作为目标标识数据进行输出。

5 综上，相对于现有方案中趋势图的绘制过程，由于本申请实施例采用简单的检索逻辑进行符合特定趋势的目标标识数据（例如交易额持续增长的用户标识或网站标识）的检索，故能够大大提高趋势判定的效率和目标标识数据的获取效率，且能够节省系统资源。

另外，相对于现有方案中分析人员在查看和分析绘制得到的趋势图后，
10 才能得到对应的趋势判定结果，本申请实施例能够直接输出符合特定趋势的目标标识数据，因此能够进一步提高趋势判定的效率和目标标识数据的获取效率，且能够节省对应的人力成本。

本申请实施例提供的方案流程可以应用于大数据的场景中，其中，上述大数据可以属于金融、电子商务等交易型数据为主的行业，也可以属于
15 医药、气象等需要存储大量的监测信息的行业，可以理解，本申请实施例对于具体的大数据不加以限制。

另外，本申请提供的方案流程可以在信息设备上运行，在实际应用中，上述信息设备具体包括但不限于：智能手机、平板电脑、电子书阅读器、MP3（动态影像专家压缩标准音频层面 3，Moving Picture Experts
20 Group Audio Layer III）播放器、MP4（动态影像专家压缩标准音频层面 4，Moving Picture Experts Group Audio Layer IV）播放器、膝上型便携计算机、车载电脑、台式计算机、机顶盒、智能电视机、可穿戴设备等等。

存储方法实施例

25 参照图 1，示出了本申请的一种数据存储方法实施例的步骤流程图，具体可以包括如下步骤：

步骤 101、依据标识数据、时间数据和对象数据的当前值之间的映射关系，确定所述对象数据在当前时间数据对应的第二时间区间内的均值，作为对应的历史均值；所述第二时间区间具体可以包括：早于所述当前时间数据的时间数据；

5 步骤 102、依据所述历史均值，建立并保存标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系。

本申请实施例中，标识数据、时间数据和对象数据的当前值之间的映射关系具体可以包括：Key（标识数据）、Time（时间数据）和 Data（对象数据的当前值）3 个字段。

10 上述标识数据可用于标识待检索的对象数据，例如，其可以包括：用户 ID（标识，Identity）、网站 ID、商品 ID 等。通常，Key 字段的值可以表示为 key_k ，其中，k 可以表示 key 的编号。

上述时间数据可用于表示对象数据的生产时间，其中，上述对象数据可以具有特定的生产周期，通常相邻的时间数据之间相隔一个生产周期。

15 例如，上述生产周期可以包括：年、月、日、小时、分钟等。通常，Time 字段的值可以表示为 $time_t$ ，其中，t 可以表示 Time 的编号。

以生产周期为日为例，若 $time_1 = "2015-01-01"$ ，则 $time_2 = "2015-01-02"$ ；

以生产周期为小时为例，若 $time_1 = "2015-01-01 01"$ ，则 $time_2 = "2015-01-01 02"$ 。

20 对象数据可用于表示待检索的数据和待确定趋势的数据，例如，上述对象数据具体可以包括：网站流量、消费金额、交易额等，其通常为整数类型。通常，Data 字段的值 $data_{k,t}$ 可以表示 key_k 在 $time_t$ 对应当前时间数据的对象数据（如访问量 20000 等）。

参照表 1（c_data），示出了本申请的一种 Key、Time 和 Data 之间的映射关系的示意，其中，Key 具体为网站 id，Time 以日为生产周期，Data

25

具体为访问量。可以理解，本申请实施例对于 Key、Time 和 Data 之间的具体映射关系不加以限制。

表 1 (c_data)

网站 id (Key)	日期 (Time)	访问量 (Data)
web1	2015-01-01	23000
web2	2015-01-01	60000
web1	2015-01-02	30000
web2	2015-01-02	50000
web1	2015-01-03	35000
web2	2015-01-03	40000
web3	2015-01-03	50000
web1	2015-01-04	40000
web2	2015-01-04	30000
web3	2015-01-04	30000
web1	2015-01-05	50000
web2	2015-01-05	40000
web3	2015-01-05	50000
web1	2015-01-06	60000
web2	2015-01-06	30000
web3	2015-01-06	40000
web1	2015-01-07	70000
web2	2015-01-07	20000
web3	2015-01-07	40000
web1	2015-01-08	50000
web2	2015-01-08	10000
web3	2015-01-08	40000

本申请实施例中，类似表 1 的映射关系中存在多条数据记录，其中，每条数据记录中均包括标识数据、当前时间数据（表示当前时间点）和对象数据的当前值。

进一步，第二时间区间可以为与当前时间数据相应的历史时间区间，其中，所述第二时间区间具体可以包括：早于所述当前时间数据的时间数据；也即，上述所述第二时间区间中至少一个时间数据早于当前时间数据，由此可以保证历史均值的历史性，从而能够反映对象数据在上述第二时间区间内的总体情况。

在实际应用中，所述第一时间区间具体可以包括：起点和终点，其中，起点对应的的时间数据可以早于终点对应的的时间数据。在本申请的一种可选实施例中，所述第二时间区间的终点可以为前一时间数据。其中，前一时间数据对应的的时间可以比当前时间数据对应的的时间早一个单位周期。假设当前时间数据为 $time_t$ ，则前一时间数据可以为 $time_{t-1}$ 。在本申请的一种应用示例中，假设上述第二时间区间的长度为 n ，则上述第二时间区间可以表示为 $[time_{t-n}, time_{t-1}]$ ，也即，上述第二时间区间具体包括当前时间数据之前的 n 个生产周期的对象数据。

在本申请的一种可选实施例中，假设 $avg(k, t, n)$ 表示 key_k 在 $time_t$ 的历史均值，则可以首先获取长度为 n 的上述第二时间区间内 key_k 的对象数据，然后，对所获取的所有对象数据进行求和，最后，将求和结果与 n 的比值作为对应的历史均值。以表 1 为例， $avg(1, t, n)$ 表示 web1 的访问量在过去 n 天的历史均值， $avg(2, t, n)$ 表示 web2 的访问量在过去 n 天的历史均值。

在实际应用中，通常第二时间区间的长度 n 的取值越大，则对历史数据的表达越准确，也即，上述历史均值所反映的对象数据在上述第二时间区间内的总体情况越准确。同时，第二时间区间的长度 n 的取值越大，则上述历史均值对应的确定过程所需的计算量越大。可以理解，本领域技术

人员可以根据实际应用需求，在准确度和计算量之间取一个较为平衡的 n 值。

例如，在本申请的另一种可选实施例中，所述第二时间区间的长度可以大于等于第一时间区间的长度的二分之一，且所述第二时间区间的长度可以小于等于第一时间区间的长度。其中，上述第一时间区间可以为用于考察和检索的时间区间，其可以依据实际应用需求确定。例如，欲要确定对象数据在某个时间区间（如最近一个月）内的特定趋势，则可以将该时间区间作为第一时间区间。可以理解，上述大于等于第一时间区间的长度 f 的二分之一、且小于等于第一时间区间的长度，仅是作为本申请的第二时间区间的长度 n 的可选实施例，实际上，本领域技术人员还可以根据实际应用需求，采用其他的 n 值，如小于 $f/2$ ，或者，大于 f 等。

在本申请的再一种可选实施例中，上述步骤 102 可以将 key_k 以及当前时间数据 t 、对象数据的当前值 $data_{k,t}$ 、过去 n 天的历史均值 $avg(k, t, n)$ ，作为一条数据记录存储到对应的数据表（以下简称检索表）中。

参照表 2，示出了本申请的一种检索表（c_data_compare）的示意，其具体可以包括：Key、Time、Data 和 Avg_data（过去 n 天的历史均值）等 4 个均值，其中的 Avg_data 是利用步骤 101 针对表 1 的映射关系对 n 取值 5 得到的。

表 2（c_data_compare）

网站 id (Key)	日期 (Time)	访问量 (Data)	Avg_data
web1	2015-01-05	50000	35600
web2	2015-01-05	40000	44000
web1	2015-01-06	60000	43000
web2	2015-01-06	30000	38000
web3	2015-01-06	40000	42000

web1	2015-01-07	70000	51000
web2	2015-01-07	20000	32000
web3	2015-01-07	40000	42000
web1	2015-01-08	50000	54000
web2	2015-01-08	10000	26000
web3	2015-01-08	40000	40000

综上，本申请实施例可以将每个生产周期的对象数据的当前值、以及过去 n 个生产周期的历史均值存储在检索表的一条数据记录中，也即，上述检索表的生产周期可以与对象数据的生产周期相同，因此能够保证历史均值数据的持续性。

检索方法实施例

参照图 2，示出了本申请的一种数据检索方法实施例的步骤流程图，具体可以包括如下步骤：

10 步骤 201、接收检索请求；所述检索请求中可以包括：预置的第一时间区间；

步骤 202、将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据；所述目标数据记录中所述当前值和所述历史均值的大小关系固定；

步骤 203、输出所述目标标识数据。

图 1 所生成检索表 (c_data_compare) 中存储了标识数据 key_k 、时间数据 $time_t$ 、对象数据的当前值 $data_{k,t}$ 和历史均值 $avg(k, t, n)$ 之间的映射关系。

应用本申请实施例，当存在第一时间区间 $[time_{t1}, time_{t2}]$ (假设区间长度为 $f=t2-t1$) 内保持特定趋势的对象数据的确定需求时，可以将上述确定

需求转化为针对上述检索表的检索需求，具体地，上述检索需求对应的检索过程具体可以包括：以 key 为分组在检索表中进行检索，以得到 $[time_{t1}, time_{t2}]$ 内、当前值 $data_{k,t}$ 和历史均值 $avg(k, t, n)$ 的大小关系固定的数据记录的数目在预置范围内的目标标识数据。

- 5 需要说明的是，上述步骤 201 可以通过预置的检索接口（如输入框等）接收用户输入的检索请求，可以理解，本申请实施例对于接收检索请求的具体过程不加以限制。

10 上述步骤 203 输出所述目标标识数据具体可以包括：在当前的信息设备上将上述目标标识数据展示给用户，或者，通过当前的信息设备向其他的信息设备发送上述目标标识数据等，本申请实施例对于上述步骤 203 输出所述目标标识数据的具体过程不加以限制。

在实际应用中，所述当前值和所述历史均值的大小关系固定，具体可以包括：所述当前值大于所述历史均值，或者，所述当前值小于所述历史均值。

- 15 在本申请的一种可选实施例中，所述目标数据记录的数目在预置范围内，具体可以包括：所述目标数据记录的数目超过阈值。

在本申请的另一种可选实施例中，所述阈值可以大于等于所述第一时间区间的长度的二分之一，且所述阈值可以小于等于所述第一时间区间的长度。假设阈值为 q ，则 $f/2 \leq q \leq f$ 。在实际应用，一种标识数据在 $[time_{t1}, time_{t2}]$ 内具体包括 f 条数据记录，故上述 $f/2 \leq q \leq f$ 能够保证 f 条数据记录中的过半数据记录均符合当前值 $data_{k,t}$ 大于（或小于）历史均值 $avg(k, t, n)$ 的条件，也即过半数据记录均能呈现出特定趋势，此种情况下，将该标识数据的趋势确定为该特定趋势，能够保证趋势确定的合理性。当然，上述 $f/2 \leq q \leq f$ 只是作为 q 的可选实施例，实际上，本领域技术人员还可以根据实际应用需求，采用小于 $f/2$ 的 q 值，本申请实施例对于具体的 q

25

值不加以限制。

在本申请的再一种可选实施例中，所述历史均值具体可以包括：对象数据值在当前时间数据对应的第二时间区间内的均值，所述第二时间区间具体可以包括：早于所述当前时间数据的时间数据。由此可以保证历史均值的
5 值的历史性，从而能够反映对象数据在上述第二时间区间内的总体情况。

在本申请的一种可选实施例中，所述将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索的步骤 202，具体可以包括：

步骤 S1、将标识数据作为分组，在标识数据、时间数据、对象数据的
10 当前值和历史均值之间的映射关系进行检索，以得到所述对象数据值和所述历史均值的大小关系固定的目标数据记录；

步骤 S2、建立并保存对象数据和所述目标数据记录的数目之间的映射关系；

步骤 S3、在所述对象数据和所述目标数据记录的数目之间的映射关系
15 中进行检索，以得到目标数据记录的数目在预置范围内的目标标识数据。

在本申请的一种应用示例中，步骤 S1 可以检索 $[time_{t1}, time_{t2}]$ 内、所有的当前值 $data_{k,t}$ 大于（或小于）历史均值 $avg(k, t, n)$ 的目标数据记录。

步骤 S2 可以 Key 字段为分组，计算对应目标数据记录的数目 D_count。
可选地，可以建立并保存对象数据 Key 和目标数据记录的数目 D_count 之
20 间的映射关系，并将该映射关系保存至中间表 tmp_c_data_compare。

进一步，步骤 S3 可以确定预置范围，并在中间表 tmp_c_data_compare 中检索 D_count 在预置范围内的目标标识数据。例如，可以根据第一时间区间的长度 f 以及对检索可靠性的要求，在 q 的取值区间 $(f/2 \leq q \text{ and } q \leq f)$ 中选取一个值，并在 tmp_c_data_compare 表中检索 $D_count \geq q$ 的标识数
25 据，作为对应的目标标识数据。

需要说明的是,上述步骤 S1-步骤 S3 的分步检索只是作为可选实施例,实际上,本领域技术人员还可以根据实际应用需求,通过一步式检索实现本申请实施例的功能。

在本申请的一种应用示例中,欲要检索 $time_{t1} = '2015-01-05'$ 到 $time_{t2} = '2015-01-08'$ 之间保持增长的 key, $f=4$, 则可以确定 $q=3$ (也即 3 天的当前值大于历史均值), 则可以构造对应的查询语句, 该查询语句的查询目标可以为 key, 查询条件可以包括: 检索表 c_data_compare 中 $time \geq '2015-01-05'$ and $time \leq '2015-01-8'$ 、且 $Data > Avg_data$ 、且满足上述条件的数据记录的数目大于等于 3。

在实际应用中,本申请实施例在应用到支持 SQL (结构化查询语言, Structured Query Language) 语言的大数据场景中,可以产生比较好的技术效果,具体地,可以大大提高检索效率。当然,本申请实施例所支持的语言不限于 SQL, C 语言、JAVA 语言也在本申请实施例的保护范围内。

综上,本申请实施例具有如下优点:

第一、本申请实施例可以将每个生产周期的对象数据的当前值、以及过去 n 个生产周期的历史均值存储在检索表的一条数据记录中,也即,上述检索表的生产周期可以与对象数据的生产周期相同,不仅能够保证历史均值数据的持续性,而且能够为本申请实施例的数据检索提供充足的数据;

第二、由于本申请实施例的标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系中可以包括多种标识数据,故本申请实施例的检索可以标识数据为单位,检索上述映射关系中当前值大于 (或者小于) 历史均值的数据记录数在预置返内的目标标识数据,也即得到的检索结果包括: 目标 key 的集合; 因此,相对于现有方案中针对每种标识数据分别绘制趋势图的过程,由于本申请实施例采用简单的检索逻辑进行符合特

定趋势的目标标识数据（例如交易额持续增长的用户标识或网站标识）的检索，故能够大大提高趋势判定的效率和目标标识数据的获取效率，且能够节省系统资源。

第三、相对于现有方案中分析人员在查看和分析绘制得到的趋势图后，才能得到对应的趋势判定结果，本申请实施例能够直接输出符合特定趋势的目标标识数据，因此能够进一步提高趋势判定的效率和目标标识数据的获取效率，且能够节省对应的人力成本。

第四、本申请能够通过阈值 q 的设定调整对应的检索质量；

第五、在 $f/2 \leq q \leq f$ 时，能够保证 f 条数据记录中的过半数据记录均能呈现出特定趋势，此种情况下，将该标识数据的趋势确定为该特定趋势，能够保证趋势确定的合理性。

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

检索装置实施例

参照图 3，示出了本申请的一种数据检索装置实施例的结构框图，具体可以包括如下模块：

接收模块 301，用于接收检索请求；所述检索请求中可以包括：预置的第一时间区间；

检索模块 302，用于将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据；所述目

标数据记录中所述当前值和所述历史均值的大小关系固定；以及

输出模块 303，用于输出所述目标标识数据。

在本申请的一种可选实施例中，所述当前值和所述历史均值的大小关系固定，具体可以包括：所述当前值大于所述历史均值，或者，所述当前
5 值小于所述历史均值。

在本申请的另一种可选实施例中，所述目标数据记录的数目在预置范围内，具体可以包括：所述目标数据记录的数目超过阈值。

在本申请的再一种可选实施例中，所述阈值可以大于等于所述第一时间区间的长度的二分之一，且所述阈值可以小于等于所述第一时间区间的
10 长度。

在本申请的又一种可选实施例中，所述检索模块 302，具体可以包括：

第一检索子模块，用于将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述对象数据值和所述历史均值的大小关系固定的目标数据记录；

15 保存子模块，用于建立并保存对象数据和所述目标数据记录的数目之间的映射关系；以及

第二检索子模块，用于在所述对象数据和所述目标数据记录的数目之间的映射关系中进行检索，以得到目标数据记录的数目在预置范围内的目标标识数据。

20 在本申请的一种可选实施例中，所述历史均值具体可以包括：对象数据值在当前时间数据对应的第二时间区间内的均值，所述第二时间区间包括：早于所述当前时间数据的时间数据。

存储装置实施例

25 参照图 4，示出了本申请的一种数据存储装置实施例的结构框图，具体可以包括如下模块：

确定模块 401，用于依据标识数据、时间数据和对象数据的当前值之间的映射关系，确定所述对象数据在当前时间数据对应的第二时间区间内的均值，作为对应的历史均值；所述第二时间区间包括：早于所述当前时间数据的时间数据；以及

- 5 保存模块 402，用于依据所述历史均值，建立并保存标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系。

在本申请的一种可选实施例中，所述第二时间区间的终点可以为前一时间数据。

- 10 在本申请的另一种可选实施例中，所述第二时间区间的长度可以大于等于第一时间区间的长度的二分之一，且所述第二时间区间的长度可以小于等于第一时间区间的长度。

对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

- 15 本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

- 本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算机程序产品。因此，本申请实施例可采用完全硬件实施例、
20 完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施

的计算机程序产品的形式。

在一个典型的配置中，上述计算机设备包括一个或多个处理器 (CPU)、输入/输出接口、网络接口和内存。内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM) 和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非持续性的电脑可读媒体 (transitory media)，如调制的数据信号和载波。

本申请实施例是参照根据本申请实施例的方法、终端设备(系统)、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序

指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上，使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围

的所有变更和修改。

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

以上对本申请所提供的一种数据检索方法、一种数据检索装置、一种数据存储方法和一种数据存储装置，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

权 利 要 求 书

1、一种数据检索方法，其特征在于，包括：

接收检索请求；所述检索请求中包括：预置的第一时间区间；

将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和
5 历史均值之间的映射关系进行检索，以得到所述第一时间区间内、目标数
据记录的数目在预置范围内的目标标识数据；所述目标数据记录中所述当
前值和所述历史均值的大小关系固定；

输出所述目标标识数据。

2、根据权利要求 1 所述的方法，其特征在于，所述当前值和所述历
10 史均值的大小关系固定，包括：所述当前值大于所述历史均值，或者，所
述当前值小于所述历史均值。

3、根据权利要求 1 所述的方法，其特征在于，所述目标数据记录的
数目在预置范围内，包括：所述目标数据记录的数目超过阈值。

4、根据权利要求 3 所述的方法，其特征在于，所述阈值大于等于所述
15 第一时间区间的长度的二分之一，且所述阈值小于等于所述第一时间区
间的长度。

5、根据权利要求 1 至 4 中任一所述的方法，其特征在于，所述将标识
数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之
间的映射关系进行检索的步骤，包括：

20 将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和
历史均值之间的映射关系进行检索，以得到所述对象数据值和所述历史均
值的大小关系固定的目标数据记录；

建立并保存对象数据和所述目标数据记录的数目之间的映射关系；

在所述对象数据和所述目标数据记录的数目之间的映射关系中进行检
25 索，以得到目标数据记录的数目在预置范围内的目标标识数据。

6、根据权利要求 1 至 4 中任一所述的方法，其特征在于，所述历史均值包括：对象数据值在当前时间数据对应的第二时间区间内的均值，所述第二时间区间包括：早于所述当前时间数据的时间数据。

5 7、一种数据存储方法，其特征在于，包括：

依据标识数据、时间数据和对象数据的当前值之间的映射关系，确定所述对象数据在当前时间数据对应的第二时间区间内的均值，作为对应的历史均值；所述第二时间区间包括：早于所述当前时间数据的时间数据；

10 依据所述历史均值，建立并保存标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系。

8、根据权利要求 7 所述的方法，其特征在于，所述第二时间区间的终点为前一时间数据。

15 9、根据权利要求 7 或 8 所述的方法，其特征在于，所述第二时间区间的长度大于等于第一时间区间的长度的二分之一，且所述第二时间区间的长度小于等于第一时间区间的长度。

10、一种数据检索装置，其特征在于，包括：

接收模块，用于接收检索请求；所述检索请求中包括：预置的第一时间区间；

20 检索模块，用于将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述第一时间区间内、目标数据记录的数目在预置范围内的目标标识数据；所述目标数据记录中所述当前值和所述历史均值的大小关系固定；以及

输出模块，用于输出所述目标标识数据。

25 11、根据权利要求 10 所述的装置，其特征在于，所述当前值和所述历史均值的大小关系固定，包括：所述当前值大于所述历史均值，或者，

所述当前值小于所述历史均值。

12、根据权利要求 10 所述的装置，其特征在于，所述目标数据记录的数目在预置范围内，包括：所述目标数据记录的数目超过阈值。

13、根据权利要求 10 至 12 中任一所述的装置，其特征在于，所述检索模块，包括：

第一检索子模块，用于将标识数据作为分组，在标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系进行检索，以得到所述对象数据值和所述历史均值的大小关系固定的目标数据记录；

保存子模块，用于建立并保存对象数据和所述目标数据记录的数目之间的映射关系；以及

第二检索子模块，用于在所述对象数据和所述目标数据记录的数目之间的映射关系中进行检索，以得到目标数据记录的数目在预置范围内的目标标识数据。

14、一种数据存储装置，其特征在于，包括：

确定模块，用于依据标识数据、时间数据和对象数据的当前值之间的映射关系，确定所述对象数据在当前时间数据对应的第二时间区间内的均值，作为对应的历史均值；所述第二时间区间包括：早于所述当前时间数据的时间数据；以及

保存模块，用于依据所述历史均值，建立并保存标识数据、时间数据、对象数据的当前值和历史均值之间的映射关系。

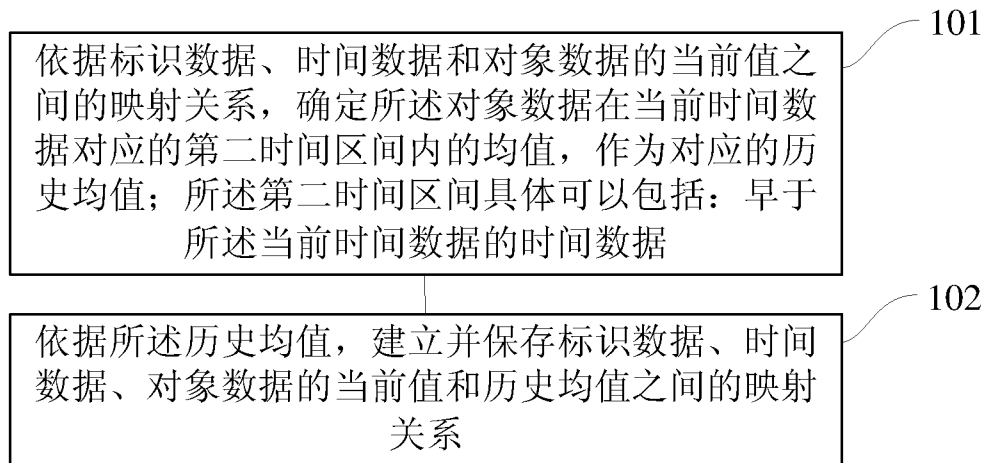


图 1

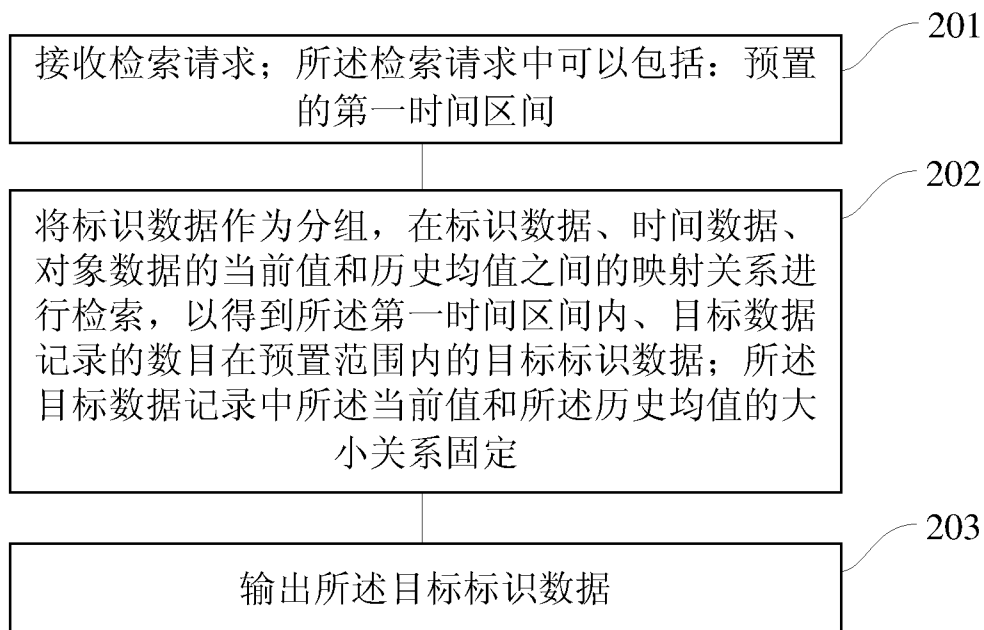


图 2

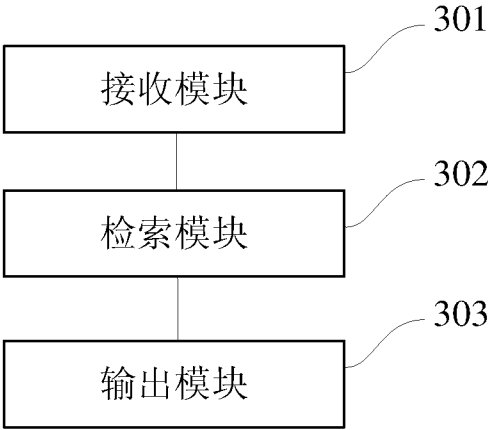


图 3

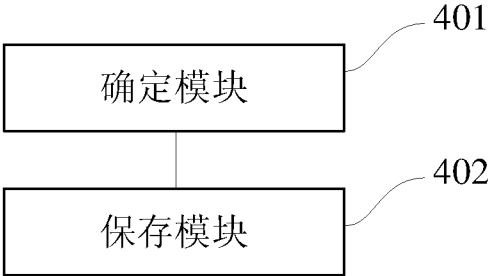


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/077450

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)

CNPAT, WPI, EPODOC, CNKI, IEEE, GOOGLE: average, history, data, statistical, time, period, query, search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104408143 A (BEIJING GRIDSUM TECHNOLOGY CO., LTD.) 11 March 2015 (11.03.2015) description, paragraphs [0007]-[0009] and [0034]-[0038]	1-14
A	CN 104462445 A (BEIJING GRIDSUM TECHNOLOGY CO., LTD.) 25 March 2015 (25.03.2015) the whole document	1-14
A	US 2005256829 A1 (ORACLE INTERNATIONAL CORPORATION) 17 November 2005 (17.11.2005) the whole document	1-14
A	US 6611726 B1 (CROSSWHITE, CARL E.) 26 August 2003 (26.08.2003) the whole document	1-14

☐ 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 14 April 2017	Date of mailing of the international search report 22 June 2017
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 KONG, Xin Telephone No. (86-10) 62413668

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/077450

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104408143 A	11 March 2015	None	
CN 104462445 A	25 March 2015	None	
US 2005256829 A1	17 November 2005	None	
US 6611726 B1	26 August 2003	None	

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)) CNPAT, WPI, EPODOC, CNKI, IEEE, GOOGLE: 均值, 历史, 数据, 统计, 时间段, 时间区间, 查询, 检索, average, history, data, statistical, time, period, query, search																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104408143 A (北京国双科技有限公司) 2015年 3月 11日 (2015 - 03 - 11) 说明书第[0007]-[0009], [0034]-[0038]段</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 104462445 A (北京国双科技有限公司) 2015年 3月 25日 (2015 - 03 - 25) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 2005256829 A1 (ORACLE INTERNATIONAL CORPORATION) 2005年 11月 17日 (2005 - 11 - 17) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 6611726 B1 (CROSSWHITE, CARL E.) 2003年 8月 26日 (2003 - 08 - 26) 全文</td> <td>1-14</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 104408143 A (北京国双科技有限公司) 2015年 3月 11日 (2015 - 03 - 11) 说明书第[0007]-[0009], [0034]-[0038]段	1-14	A	CN 104462445 A (北京国双科技有限公司) 2015年 3月 25日 (2015 - 03 - 25) 全文	1-14	A	US 2005256829 A1 (ORACLE INTERNATIONAL CORPORATION) 2005年 11月 17日 (2005 - 11 - 17) 全文	1-14	A	US 6611726 B1 (CROSSWHITE, CARL E.) 2003年 8月 26日 (2003 - 08 - 26) 全文	1-14
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 104408143 A (北京国双科技有限公司) 2015年 3月 11日 (2015 - 03 - 11) 说明书第[0007]-[0009], [0034]-[0038]段	1-14															
A	CN 104462445 A (北京国双科技有限公司) 2015年 3月 25日 (2015 - 03 - 25) 全文	1-14															
A	US 2005256829 A1 (ORACLE INTERNATIONAL CORPORATION) 2005年 11月 17日 (2005 - 11 - 17) 全文	1-14															
A	US 6611726 B1 (CROSSWHITE, CARL E.) 2003年 8月 26日 (2003 - 08 - 26) 全文	1-14															
☐ 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 2017年 4月 14日		国际检索报告邮寄日期 2017年 6月 22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 孔昕 电话号码 (86-10) 62413668															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/077450

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	104408143	A	2015年 3月 11日	无	
CN	104462445	A	2015年 3月 25日	无	
US	2005256829	A1	2005年 11月 17日	无	
US	6611726	B1	2003年 8月 26日	无	