

(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
2020 年 10 月 1 日 (01.10.2020)



(10) 国际公布号
WO 2020/192007 A1

(51) 国际专利分类号:
G06F 21/60 (2013.01)

(21) 国际申请号: PCT/CN2019/102746

(22) 国际申请日: 2019 年 8 月 27 日 (27.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910241311.X 2019年3月28日 (28.03.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

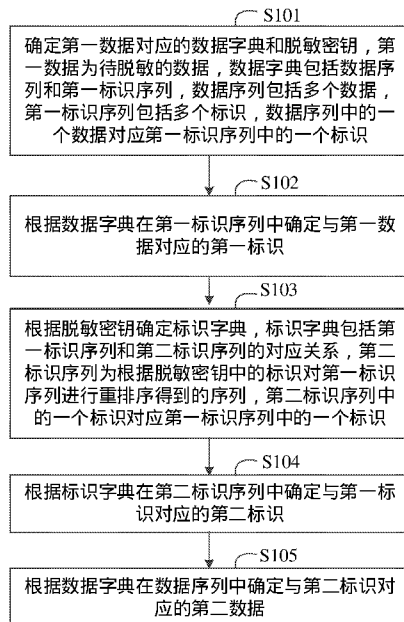
(72) 发明人: 高勇(GAO, Yong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 董晓琼(DONG, Xiaoqiong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 广州三环专利商标代理有限公司(SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路 80 号汇华商贸大厦 1508 室, Guangdong 510070 (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,

(54) Title: DATA DESENSITIZATION METHOD AND RELATED DEVICE

(54) 发明名称: 数据脱敏方法和相关装置



- S101 Determine a data dictionary and a desensitization key corresponding to first data, wherein the first data is data to be desensitized, the data dictionary comprises a data sequence and a first identifier sequence, the data sequence comprises multiple pieces of data, the first identifier sequence comprises multiple identifiers, and one piece of data in the data sequence corresponds to one identifier in the first identifier sequence
- S102 Determine, according to the data dictionary, a first identifier corresponding to the first data from the first identifier sequence
- S103 Determine an identifier dictionary according to the desensitization key, wherein the identifier dictionary comprises a correlation between the first identifier sequence and a second identifier sequence, the second identifier sequence is a sequence obtained by reordering the first identifier sequence according to an identifier in the desensitization key, and an identifier in the second identifier sequence corresponds to an identifier in the first identifier sequence
- S104 Determine, according to the identifier dictionary, a second identifier corresponding to the first identifier from the second identifier sequence
- S105 Determine, according to the data dictionary, second data corresponding to the second identifier from the data sequence

图 1

(57) Abstract: Provided are a data desensitization method and a related device. The method comprises: determining a data dictionary and a desensitization key corresponding to first data, wherein the data dictionary comprises a correlation between a data sequence and a first identifier sequence, and data in the data sequence corresponds to an identifier in the first identifier sequence; determining, according to the data dictionary, a first identifier corresponding to the first data; determining an identifier dictionary according to the desensitization key, wherein the identifier dictionary comprises a correlation between the first identifier sequence and a second

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)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

identifier sequence, the second identifier sequence is a sequence obtained by reordering the first identifier sequence, and an identifier in the second identifier sequence corresponds to an identifier in the first identifier sequence; determining, according to the identifier dictionary, a second identifier corresponding to the first identifier from the second identifier sequence; and determining, according to the data dictionary, second data corresponding to the second identifier, and determining the second data to be desensitized data corresponding to the first data. According to the method, the security of data that is not desensitized and the reducibility of desensitized data can be taken into consideration at the same time.

(57) 摘要: 一种数据脱敏方法和相关装置, 其中, 方法包括: 确定第一数据对应的数据字典和脱敏密钥, 数据字典包括数据序列和第一标识序列的对应关系, 数据序列中的数据对应第一标识序列中的标识; 根据数据字典确定与第一数据对应的第一标识; 根据脱敏密钥确定标识字典, 标识字典包括第一标识序列和第二标识序列的对应关系, 第二标识序列为对第一标识序列进行重排序得到的序列, 第二标识序列中的标识对应第一标识序列中的标识; 根据标识字典在第二标识序列中确定与第一标识对应的第二标识; 根据数据字典确定与第二标识对应的第二数据, 并将第二数据确定为第一数据对应的脱敏数据。该方法可同时兼顾未脱敏数据的安全性和脱敏数据的可还原性。

数据脱敏方法和相关装置

本申请要求于2019年03月28日提交中国专利局、申请号为201910241311X、申请名称为“数据脱敏方法和相关装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及信息安全技术领域，尤其数据脱敏方法和相关装置。

背景技术

数据脱敏，又称数据漂白、数据去隐私化或数据变形，是指对某些敏感信息按照一定规则进行数据的变形，使得数据变换为另一个数据，以实现对敏感隐私数据的可靠保护。在涉及客户安全数据或者一些商业性敏感数据的场景下，在不违反系统规则条件下，在利用真实数据进行改造并提供测试使用时需要对如身份证号、手机号、卡号、客户号等个人信息进行数据脱敏后才能使用。

数据脱敏按脱敏规则划分可分为可恢复脱敏与不可恢复脱敏。其中，可恢复脱敏是指脱敏后的数据通过一定的方式，可以恢复成原来的敏感数据；不可恢复脱敏是指脱敏后的数据被脱敏的部分使用任何方式都不能恢复出。目前，数据脱敏一般是通过建立一个数据字典，然后根据该数据字典进行一定的映射，得到脱敏后的数据。一种方式是根据数据字典随机映射，这种方式即为不可恢复脱敏，但存在的问题在于进行随机脱敏后的数据无法还原为原始数据；另一种方式是根据数据字典按照一定的映射规则进行映射，这种方式为可恢复脱敏，但存在的问题在于映射关系容易被获知的风险，安全性不足。

申请内容

本申请实施例提供数据脱敏方法和相关装置，解决目前的脱敏方法无法兼顾数据还原和数据安全的问题。

第一方面，提供一种数据脱敏方法，包括：

确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个标识对应所述第一标识序列中的一个数据，所述数据脱敏密钥包括至少一个所述第一标识序列中的标识；

根据所述数据字典在所述第一标识序列中确定与所述第一数据对应的第一标识；

根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为根据所述脱敏密钥中的标识对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

根据所述数据字典确定在所述数据序列中与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

第二方面，提供一种数据脱敏装置，包括：

字典密钥确定模块，用于确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个数据对应所述第一标识序列中的一个标识，所述脱敏密钥包括至少一个所述第一标识序列中的标识；

第一标识确定模块，用于根据所述数据字典确定在所述第一标识序列中确定与所述第一数据对应的第一标识；

标识字典确定模块，用于根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为根据所述脱敏密钥中的标识对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

第二标识确定模块，用于根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

脱敏数据确定模块，用于根据所述数据字典在所述数据序列中确定与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

第三方面，提供另一种数据脱敏装置，包括处理器、存储器以及输入输出接口，所述处理器、存储器和输入输出接口相互连接，其中，所述输入输出接口用于输入或输出数据，所述存储器用于存储数据脱敏装置执行上述方法的应用程序代码，所述处理器被配置用于执行上述第一方面的方法。

第四方面，提供一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行上述第一方面的方法。

通过本申请实施例可以同时兼顾数据的安全性和脱敏数据的可还原性。

附图说明

图1是本申请实施例提供的一种数据脱敏方法的流程示意图；

图2是本申请实施例提供的一种数据与标识的变换的示意图；

图3是本申请实施例提供的一种映射示意图；

图4是本申请实施例提供的一种数据脱敏装置的组成结构示意图；

图5是本申请实施例提供的另一种数据脱敏装置的组成结构示意图。

具体实施方式

下面将结合图1至图5，对本申请实施例提供的**数据脱敏**方法和相关装置进行说明。

本申请实施例的方案可以应用于金融公司、政府单位、医院、银行等各个行业中需要利用敏感数据进行数据挖掘、数据分析等使用敏感数据的场景中。在这些场景中，在利用敏感数据进行数据挖掘或数据分析之前，需要对这些敏感数据进行数据脱敏，以防止敏感数据泄露。例如，本申请实施例的方案具体可应用于社保数据中心需要将社保数据提供给外包开发人员进行社保系统的开发和测试，在这种场景下，需要将社保数据进行数据脱敏后提供给外包开发人员。又如，本申请实施例的方案具体还可应用于运营商将客户信息、通话记录等业务数据提供给如监察机构、公安机关、政府部门等不同身份的访问者，运营

商需要根据不同身份的用户配置不同的数据脱敏规则以满足其不同的需求。本申请实施例中，敏感数据包括但不限于姓名、身份证号码、居住地址、电话号码、银行账号、邮箱地址、所属城市、邮编、各种密码（如账户查询密码、取款密码、登录密码等）、组织机构名称、营业执照号码、银行账号、交易日期、交易金额等于用户主体或用户主体的行为有关的数据。

以下介绍申请实施例的技术方案。

参见图 1，图 1 是本申请实施例提供的一种数据脱敏方法的流程示意图，如图所示，该方法包括如下步骤：

S101，确定第一数据对应的数据字典和脱敏密钥，第一数据为待脱敏的数据，数据字典包括数据序列和第一标识序列，数据序列包括多个数据，第一标识序列包括多个标识，数据序列中的一个数据对应第一标识序列中的一个标识。

这里，第一数据对应的数据字典为与第一数据所属的类型或类别相对应的数据字典，数据序列包括该第一数据所属的类型或类别对应的所有可能的数据。其中，第一数据所属的类型或类别具体是指当前待脱敏的数据具体是什么类型或类别的数据。例如，当前待脱敏的数据为身份证号码，则第一数据所属的类型或类别为身份证号码，数据序列包括身份证号码对应的所有可能的数据，即第一数据字典包括 0 至 9 中的每一个数字以及 X。又如，当前待脱敏的数据为邮箱地址，则第一数据所属的类型或类别为邮箱地址，数据序列包括邮箱地址对应的所有可能的数据，即数据序列包括 0 至 9 中的每一个数字、A 至 Z 和 a 至 z 中的每一个英文字母以及邮箱地址可能用到的其他字符，如“.”、“/”等字符。

这里，第一标识序列为根据第一数据序列中的数据的个数和类型所设置的个数与该数据序列中的数据的个数相同的，并且，类型与第一数据序列中的数据的类型相同或不同的标识的组合。例如，数据序列包括身份证号码对应的所有可能的数据，即第一数据字典包括 0 至 9 中的每一个数字以及 X，则第一标识序列可以为由英文字母 a 至 k 组成的序列，第一标识序列也可以为数字 0 至 9 以及 X 组成的序列。本申请实施例中，第一标识序列中的标识可以包括与脱敏任务相关的和/或与脱敏操作人员相关的所有的可能的标识。例如，第一标识序列中的标识包括脱敏任务的编号对应的所有可能的标识，假设脱敏任务的编号由 0 至 9 中的任一数字以及 A 和 B 两个字母组成，则第一标识序列中的标识可以包括 0 至 9 中的每一个数字以及 A 和 B 两个字母。又如，第一标识序列中的标识也包括脱敏操作人员的工号对应的所有可能的标识，假设脱敏操作人员的工号由 0 至 9 中的任一数字组成，则第一标识序列中的标识可以包括 0 至 9 中的每一个数字。又如，第一标识序列中的标识序列包括脱敏任务的编号对应的所有可能的标识和脱敏操作人员的工号对应的所有可能的标识，假设脱敏任务的编号由 0 至 9 中的任一数字以及 A 和 B 两个字母组成，脱敏操作人员的工号由 0 至 9 中的任一数字组成，则第一标识序列中的数据标识可以包括 0 至 9 中的每一个数据以及 A 和 B 两个字母。不限于这里的描述。

这里，脱敏密钥由第一标识序列中的至少一个标识组成。在一种可能的实施方式中，该脱敏密钥可以为与脱敏任务相关的某一个账号或者号码或者标识或编号。例如，脱敏密钥可以为脱敏任务的编号，或者，脱敏密钥可以为执行当前脱敏任务的脱敏操作人员的工号、脱敏编号，或者，第一脱敏密钥可以为脱敏任务的编号与执行当前脱敏任务的脱

敏操作人员的工号、脱敏编号。

以下举例来对第一数据、第一数据对应的数据字典和脱敏密钥进行说明。例如，当前执行的脱敏任务为对用户的电话号码进行数据脱敏，脱敏任务的编号为 0x001，当前执行脱敏任务的脱敏操作人员的工号为 bj856，脱敏任务编号的格式为“0x+数字编号”，脱敏操作人员的工号格式为“bj+数字编号”，电话号码由数字组成，则第一数据为用户的电话号码中的一个数据，第一数据对应的数据字典中如图 2 中的 A 所示，其中，图 2 中 A 的左边一列（数据）为数据序列，第一数据序列由电话号码所有可能的数字组成，右边一列（标识 1）为第一标识序列，第一标识序列由脱敏任务编号中的数字编号的所有可能的数字或者脱敏操作人员的工号中的数字编号的所有可能的数字组成。第一脱敏密钥可以为“001”或“856”，当第一脱敏密钥为“001”时，第一脱敏密钥与脱敏任务的脱敏任务编号相关，第一脱敏密钥可以用于表示脱敏任务的任务编号；当第一脱敏密钥为“856”时，第一脱敏密钥与当前执行脱敏任务的脱敏操作人员的工号相关，其第一脱敏密钥可以用于表示当前执行脱敏任务的脱敏操作人员的工号。

具体实现中，可以针对脱敏场景中涉及的数据类别预设一个或多个数据字典，使得该一个或多个数据字典中的第一数据序列中的数据能够涵盖该数据类别可能存在的所有数据，并保存数据类别与数据字典之间的对应关系。在进行数据脱敏时，可根据第一数据所属的类别或类型确定与第一数据对应的数据字典，然后根据该第一数据在当前待脱敏的所有数据中的排列顺序确定第一数据对应的第一脱敏密钥。

S102，根据数据字典在第一标识序列中确定与第一数据对应的第一标识。

例如，数据字典如图 2 中的 A 所示，第一数据为 3，可根据数据字典确定在第一标识序列（标识 1）中确定 3 为与数据“3”对应的第一标识。

S103，根据脱敏密钥确定标识字典，标识字典包括第一标识序列和第二标识序列的对应关系，第二标识序列为根据脱敏密钥中的标识对第一标识序列进行重排序得到的序列，第二标识序列中的一个标识对应第一标识序列中的一个标识。

这里，对第一标识序列中的标识进行重排序是指将第一标识序列中的标识在第一标识序列中的顺序进行变化和调整，使得这些标识在变化后的标识序列中的排列顺序与在第一标识序列中的排列顺序不同，即对第一标识序列进行重排序得到的序列中包含的标识与第一标识序列中包含的标识的内容和个数相同，标识在对第一标识序列进行重排序得到的序列中的顺序和标识在第一标识序列中的顺序不同。

例如，第一标识序列如图 2 中的 A 的右边一列所示，则对第一标识序列中的标识进行重排序后的标识序列可以如图 2 中的 B 所示。对比图 2 中的 B 和 A 可知，B 中的标识的排列顺序相对于 A 中的标识的排列顺序发生了变化。

本申请实施例中，标识字典可以由脱敏密钥对应的一个或多个子标识字典组成，脱敏密钥对应的子标识字典的个数可以与脱敏密钥包含的标识的个数有关。以下具体介绍根据脱敏密钥确定标识字典的方式。

根据脱敏密钥确定标识字典可以分为以下两个步骤：

步骤一、确定脱敏密钥中的各个标识对应的子标识字典。以脱敏密钥包括 m 个标识为例，m 为正整数，m 大于或等于 1。由于脱敏密钥包括 m 个标识，则根据脱敏密钥中的 m

个标识确定的子标识字典有 m 个, 即首先确定 m 个标识各自对应的 m 个子标识字典。

在一种可行的实施方式中, 可以根据预设的重排序规则和脱敏密钥中的 m 个标识分别对第一标识序列进行重排序, 得到脱敏密钥中的 m 个标识各自对应的 m 个重排序标识序列, 然后将第一标识序列分别与脱敏密钥中的 m 个标识各自对应的 m 个重排序标识序列组成脱敏密钥中的 m 个标识各自对应的 m 个子标识字典。

具体地, 预设的重排序规则可以有多种。在一种具体实现方式中, 预设的重排序规则可以为对第一标识序列进行按顺序移位, 使得目标标识在移位得到的重排序标识序列中处于第 k 位, k 为大于或等于 1 并且小于或等于 N 的任一正整数, N 为第一标识序列中标识的个数。这里, 目标标识是要确定与其相对应的重排序标识序列的标识。

举例进行说明, 例如, 脱敏密钥为 856, 第一标识序列如图 2 中的 A 的右边一列数据所示, 则根据预设的重排序规则和脱敏密钥中的 m 个标识分别对第一标识序列进行重排序, 得到脱敏密钥中的 m 个标识各自对应的 m 个重排序标识序列的过程如下:

1、将第一标识序列中的标识进行按顺序移位, 使得脱敏密钥中的标识“8”在移位后的序列中位于第 1 位, 移位后的序列如图 2 中的 B 所示, 图 2 中的 B 所示的序列即为标识“8”对应的重排序标识序列。

2、将第一标识序列中的标识进行按顺序移位, 使得脱敏密钥中的标识“5”在移位后的序列中位于第 1 位, 移位后的序列如图 2 中的 C 所示, 图 2 中的 C 所示的序列即为标识“5”对应的重排序标识序列。

3、将第一标识序列中的标识进行按顺序移位, 使得脱敏密钥中的标识“6”在移位后的序列中位于第 1 位, 移位后的序列如图 2 中的 D 所示, 图 2 中的 D 所示的序列即为标识“6”对应的重排序标识序列。

进一步地, 将第一标识序列分别与脱敏密钥中的 m 个标识各自对应的 m 个重排序标识序列组成脱敏密钥中的 m 个标识各自对应的 m 个子标识字典的过程可以如下:

1、将第一标识序列与标识“8”对应的重排序标识序列组成标识“8”对应的子标识字典, “8”对应的子标识字典如图 2 中的 E 所示。

2、将第一标识序列与标识“5”对应的重排序标识序列组成标识“5”对应的子标识字典, “5”对应的子标识字典如图 2 中的 F 所示。

3、将第一标识序列与标识“6”对应的重排序标识序列组成标识“6”对应的子标识字典, “6”对应的子标识字典如图 2 中的 G 所示。

在另一种具体实现方式中, 重排序规则还可以为在第一标识序列中将目标标识与位于第 h 位的标识进行替换, 得到目标标识对应的重排序标识序列, h 为大于或等于 1 并且小于或等于 N 的任一正整数。

在另一种可行的实施方式中, 也可以事先根据预设的重排序规则和第一标识序列中的 N 个标识为第一标识序列中的 N 个标识设置与其相对应的 N 个重排序标识序列, 然后分别建立 N 个重排序标识序列与第一标识序列中的 N 个标识之间的对应关系, 其中, N 为第一标识序列中的标识的总个数。在根据脱敏密钥确定标识字典时, 根据 N 个重排序标识序列与 N 个标识之间的对应关系, 以及, 脱敏密钥中的 m 个标识, 确定与脱敏密钥中的 m 个标识相对应的 m 个重排序标识序列, 然后将第一标识序列分别与脱敏密钥中的 m 个标识对

应的重排序标识序列组成 m 个子标识字典。

步骤二、按照预设的排列组合规则将 m 个标识各自对应的 m 个子标识字典进行可重复性地排列和组合，以得到标识字典。将 m 个子标识字典进行可重复性地排列和组合所得到的标识字典包括 M 个子标识字典， M 为正整数， M 大于或等于 m 。该 M 个子标识字典中的第 i 个子标识字典包括第一标识序列和第 i 标识序列，该第 i 标识序列为该第 i 个子标识字典对应的标识所对应的重排序标识序列， i 为正整数， i 大于或等于 1 并且小于或等于 M 。例如，第 i 个子标识字典为脱敏密钥中的第 1 个标识对应的子标识字典，则第 i 标识序列为脱敏密钥中的第 1 个标识对应的重排序标识序列。其中，第 i 标识序列中的第 M 标识序列为第二标识序列。

这里，可重复性地排列和组合，是指在按照预设的排列组合规则对 m 个子标识字典进行排列组合的过程中，可多次或重复使用 m 个子标识字典中的一个或多个子标识字典与 m 个子标识字典中的其他标识字典进行排列组合。例如，得到 4 个子标识字典，分别为子标识字典 1、子标识字典 2、子标识字典 3 以及子标识字典 4，则在按照预设的排列组合规则对 m 个子标识字典进行排列组合的过程中，使用子标识字典 1 两次，与子标识字典 2、子标识字典 3 以及子标识字典 4 进行排列组合，以此得到包含 5 个子标识字典的标识字典。

具体地，预设的排列组合规则可以有多种。在一种可能的实施方式中，排列组合规则可以为：按脱敏密钥中的标识在脱敏密钥中的排列顺序对子标识字典进行排列组合，使得排列组合得到的标识字典包含的子标识字典的个数等于脱敏密钥中的标识的个数，以及，进行排列组合后的各个子标识字典在标识字典中的顺序等于各个子标识字典对应的标识在脱敏密钥中的顺序。即使得 $M=m$ ，以及第 i 个子标识字典为根据脱敏密钥中的第 j 个标识确定的子标识字典， $j=i$ ， j 为大于或等于 1 并且小于或等于 m 的正整数。

举例来进行说明，例如，脱敏密钥中的标识分别为标识 1、标识 2、标识 3，标识 1 在脱敏密钥中的顺序为 1，标识 2 在脱敏密钥中的顺序为 2，在脱敏密钥中的顺序为 3，通过步骤一确定标识 1 对应的子标识字典为子标识字典 1，通过步骤一确定标识 2 对应的子标识字典为子标识字典 2，通过步骤一确定标识 3 对应的子标识字典为子标识字典 3。按照上述预设的排列组合规则将子标识字典进行排列组合得到的标识字典包括子标识字典 1、标识字典 2 以及子标识字典 3，并且，子标识字典 1 在标识字典中的顺序为 1，标识字典 2 在标识字典中的顺序为 2，子标识字典 3 在标识字典中的顺序为 3。

在另一种可能的实施方式中，排列组合规则还可以为：按脱敏密钥中的标识在脱敏密钥中的排列顺序对子标识字典进行排列组合，使得排列组合得到的标识字典包含的子标识字典的个数等于脱敏密钥中的标识的个数的两倍减一，进行排列组合后的各个子标识字典在标识字典中的顺序等于各个子标识字典对应的标识在脱敏密钥中的顺序，或者，等于各个子标识字典在标识中的逆向顺序等于各个子标识字典对应的标识。即使得 $M=2m-1$ ，第 i 个子标识字典为根据脱敏密钥中的第 j 个标识确定的子标识字典， $j = \begin{cases} i, i \leq m \\ (i+1)/2, i > m \end{cases}$ ， j 为大于或等于 1 并且小于或等于 m 的正整数。

举例来进行说明，例如，脱敏密钥中的标识分别为标识 1、标识 2、标识 3，标识 1 在脱敏密钥中的顺序为 1，标识 2 在脱敏密钥中的顺序为 2，在脱敏密钥中的顺序为 3，通过

步骤一确定标识 1 对应的子标识字典为子标识字典 1, 通过步骤一确定标识 2 对应的子标识字典为子标识字典 2, 通过步骤一确定标识 3 对应的子标识字典为子标识字典 3。那么, 按照上述排列组合规则将子标识字典进行排列组合得到的标识字典包括 2 个子标识字典 1、2 个标识字典 2 以及 1 个子标识字典 3, 并且, 子标识字典 1 在标识字典中的顺序为 1 和 5, 标识字典 2 在标识字典中的顺序为 2 和 4, 子标识字典 3 在标识字典中的顺序为 3。

不限于上述两种排列组合规则, 在可选实施方式中, 还可以有其他的排列组合规则, 本申请实施例不做限制。

S104, 根据标识字典在第二标识序列中确定与第一标识对应的第二标识。

这里, 在脱敏密钥包括一个标识的情况下, 标识字典由第一标识序列和第二标识序列组成, 根据标识字典在第二标识序列中确定第一标识对应的第二标识的具体实现方式为: 将第二标识序列中与第一标识对应的标识确定为第二标识。

举例来说明, 例如, 标识字典由图 2 中 B 和图 2 中的 A 中的第一标识序列 (标识 1) 组成, 则标识字典可以如图 2 中的 E 所示。假设确定的第一标识为 3, 则根据 E 确定的第二标识为 1。

这里, 在脱敏密钥包括多个标识的情况下, 标识字典由 M 个子标识字典组成, M 个子标识字典中的第 i 个子标识字典包括第一标识序列和第 i 标识序列, 第 i 标识序列为对第一标识序列进行重排序得到的序列, 第 i 标识序列中的一个标识对应第一标识序列中的一个标识, 第 M 标识序列为第二标识序列。根据标识字典在第二标识序列中确定第一标识对应的第二标识的具体实现方式为: 根据第 i 个子标识字典在第 i 标识序列中确定与第 i 标识对应的第 (i+1) 标识, 其中, 当 i=1 时, 第 i 标识为第一标识, 将第 M 标识序列中的第 (M+1) 标识确定为第二标识。其具体过程可以为: 根据第 1 个子标识字典在第 1 标识序列中确定与第一标识对应的第 2 标识; 根据第 2 个子标识字典在第 2 标识序列中确定与第 2 标识对应的第 3 标识;; 根据第 M 个子标识字典在第 M 标识序列中确定与第 M 标识对应的第 (M+1) 标识。

举例来进行说明, M=5, 第一标识为 3, 子标识字典分别如图 2 中的 E、F、G 所示。其中, E 所示的子标识字典在标识字典中的顺序为 1 和 5, F 所示的子标识字典在标识字典中的顺序为 2 和 4, G 所示的子标识字典的在标识字典中的顺序为 3, 则上述过程可以如图 3 所示, 首先根据第 1 个子标识字典在第 1 标识序列中与第一标识对应的第 2 标识, 第 1 个子标识字典为 E 所示的子标识字典, 第 1 标识序列为 E 所示的子标识字典中标识 2 这一标识序列, 即根据 E 所示的子标识字典确定在标识 2 这一标识序列中确定 3 对应的标识, 3 对应的标识为 1; 然后根据第 2 个子标识字典在第 2 标识序列中与第 2 标识对应的第 3 标识, 第 2 个子标识字典为 F 所示的子标识字典, 第 2 标识序列为 F 所示的子标识字典中标识 2 这一标识序列, 即根据 F 所示的子标识字典在标识 2 这一标识序列中确定 1 对应的标识, 1 对应的标识为 6; 接着, 根据第 3 个子标识字典在第 3 标识序列中与第 3 标识对应的第 4 标识, 第 3 个子标识字典为 G 所示的子标识字典, 第 1 标识序列为 G 所示的子标识字典中标识 2 这一标识序列, 即根据 G 所示的子标识字典在标识 2 这一标识序列中确定 6 对

应的标识, 6 对应的标识为 2; 接着, 根据第 4 个子标识字典在第 4 标识序列中与第 4 标识对应的第 5 标识, 第 4 个子标识字典为 F 所示的子标识字典, 第 4 标识序列为 F 所示的子标识字典中标识 2 这一标识序列, 即根据 F 所示的标识字典确定 2 对应的标识, 2 对应的标识为 7; 最后, 根据第 5 个子标识字典在第 5 标识序列中与第 5 标识对应的第 6 标识, 第 5 个子标识字典为 E 所示的子标识字典, 第 1 标识序列为 E 所示的子标识字典中标识 2 这一标识序列, 即根据 E 所示的标识字典确定 7 对应的标识, 7 对应的标识为 5。

S105, 根据数据字典在数据序列中确定与第二标识对应的第二数据。

例如, 数据字典如图 2 中的 A 所示, 根据标识字典确定的第二标识为 7, 则根据数据字典在数据序列 (数据) 中确定标识 “7” 对应的数据, 将标识 “7” 对应的数据确定为第二数据, 标识 “7” 对应的数据为 7, 则将数据 “7” 确定为第一数据对应的脱敏数据。

结合图 2 和图 3 可知, 通过上述方案, 将数据 3 变换成了数据 7, 实现了对数据的脱敏。在上述方案中, 根据数据脱敏密钥确定标识字典, 然后利用数据字典和标识字典对数据进行多次映射得到脱敏后的数据, 在标识字典中对标识的映射相当于打乱了数据字典原始的映射顺序, 使得映射具备随机映射的特性, 其他人根据脱敏后的数据无法推导出映射关系, 保证了数据的安全性, 由于是通过数据脱敏密钥中的标识得到标识字典, 那么根据脱敏密钥可确定标识字典, 然后根据逆向的映射过程, 可以对脱敏后的数据进行还原, 保证了脱敏后的数据的可还原性, 即, 本申请实施例可同时兼顾数据的安全性和脱敏数据的可还原性。

在一种可行的实施方式中, 可以对待脱敏的属于同一类型的数据均按上述图 1 所示的方法进行数据脱敏。

在另一种可行的实施方式中, 也可以采用不同的脱敏密钥对待脱敏的属于同一数据整体的不同数据进行数据脱敏, 数据整体中的两个相邻数据所对应的脱敏密钥之间可以相互关联。其中, 数据整体可以是指完整的组合起来拥有实际含义的数据的集合。具体地, 数据整体可以为一个姓名、一个身份证号码、一个居住地址、一个电话号码、一个银行账号或一个邮箱地址等对应的所有数据。例如, 数据整体为一个身份证号码对应的所有数据, 对于中国大陆的身份证号码来说, 该数据整体为一个包含 18 位数字的数据集合。可选地, 数据整体也可以是指待脱敏的所有数据的集合, 例如, 待脱敏的数据一共有 100 个, 那么数据整体可以是指包含了这 100 个数据的集合。

可选地, 上述方法还可以包括如下步骤, 确定标识字典中的第 A 个子标识字典, 第 A 个子标识字典为根据脱敏密钥中的第 a 个标识确定的子标识字典,

$$a = \begin{cases} m, B = n * (N * m) \\ [B/N] \% m, B \neq a * (N * m) \end{cases}, \text{ B 为第一数据在待脱敏的数据整体中的排列顺序, 待脱}$$

敏的数据整体包括至少两个数据, a 和 n 均为正整数, N 等于第一标识序列中的标识的个数; 根据第 A 个子标识字典确定第 A 标识序列中的第三标识, 第三标识在第 A 标识序列中位于

第 a 个标识的后一位；在脱敏密钥中将第 a 个标识替换为第三标识；将替换得到的脱敏密钥确定为第一数据的下一个数据对应的脱敏密钥。

举例来进行说明，例如，待脱敏的数据整体为一个身份证号码对应的所有数据的集合，那么数据字典的第一数据序列包括 0 至 9 以及 X 这 11 个数据，即 $N=11$ ，假设待脱敏的数据整体为 5003101111111111111，第一数据为 3，数据脱敏密钥为 856，3 在待脱敏的数据整体中的排列顺序为 4，则确定 $B=4$ ， $m=3$ ， $N*m$ 为 33， B 不为 $N*m$ 的整数倍， $a=[4/11]\%3=1$ ，那么在标识字典中确定根据脱敏密钥中的第 1 个标识确定的子标识字典，即确定标识“8”对应的子标识字典，假设标识 8 对应的标识字典为如图 2 中的 E 所示，图 2 中的 E 所示的第二标识序列（标识 2）中排列在 8 下一位的为 9，则将 9 确定为第三标识，然后在第一脱敏密钥中将 8 替换为 9，替换得到的脱敏密钥为 956，将 956 确定为 3 的下一个数据“1”对应的脱敏密钥。

针对同一个数据整体中的不同数据，采用不同的脱敏密钥，可进一步保证数据的安全性，同时根据相邻数据对应的脱敏密钥之间的联系，根据其中一个脱敏密钥可推得其他的脱敏密钥，从而可以分别对数据进行还原，也能保证数据的可还原性。

上面介绍了申请实施例的方法，下面介绍申请实施例的装置。

参见图 4，图 4 是本申请实施例提供的一种数据脱敏装置的组成结构示意图，该装置 30 包括：

字典密钥确定模块 301，用于确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个数据对应所述第一标识序列中的一个标识，所述脱敏密钥包括至少一个所述第一标识序列中的标识；

第一标识确定模块 302，用于根据所述数据字典确定在所述第一标识序列中确定与所述第一数据对应的第一标识；

标识字典确定模块 303，用于根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为根据所述脱敏密钥中的标识对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

第二标识确定模块 304，用于根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

脱敏数据确定模块 305，用于根据所述数据字典在所述数据序列中确定与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

在一种可能的设计中，所述脱敏密钥包括 m 个标识，所述标识字典包括 M 个子标识字典，所述 M 个子标识字典由可重复排列的 m 个子标识字典排列组合得到，所述 m 个子标识字典为根据所述 m 个标识字典确定的子标识字典， m 、 M 均为正整数， M 大于或等于 m ；

所述 M 个子标识字典中的第 i 个子标识字典包括所述第一标识序列和第 i 标识序列的

对应关系, 所述第 i 标识序列为对所述第一标识序列进行重排序得到的序列, 所述第 i 标识序列中的一个标识对应所述第一标识序列中的一个标识, 第 M 标识序列为所述第二标识序列, i 为正整数, i 大于或等于 1 并且小于或等于 M ;

第二标识确定模块 303 具体用于:

根据所述第 i 个子标识字典在所述第 i 标识序列中确定与所述第 i 标识对应的第 $(i+1)$ 标识, 其中, 当 $i=1$ 时, 所述第 i 标识为所述第一标识;

将第 M 标识序列中的第 M 标识确定为所述第二标识。

在一种可能的设计中, 所述标识字典确定模块 303, 具体用于:

确定所述脱敏密钥中的各个标识对应的子标识字典;

按照预设的排列组合规则, 分别将所述各个标识各自对应的子标识字典进行可重复性地排列和组合。

在一种可能的设计中, $M=m$, 所述第 i 个子标识字典为根据所述脱敏密钥中的第 j 个标识确定的子标识字典, $j=i$, j 为正整数, j 大于或等于 1 并且小于或等于 m 。

在一种可能的设计中, $M=2m-1$, 所述第 i 个子标识字典为根据所述脱敏密钥中的第 j 个标识确定的子标识字典, $j = \begin{cases} i, i \leq m \\ (i+1)/2, i > m \end{cases}$ 。

在一种可能的设计中, 所述第 i 标识序列为对所述第一标识序列进行按顺序移位得到的序列, 并且, 所述第 j 个标识在所述第 i 个子标识字典的第 i 标识序列中位于第 k 位, k 为大于或等于 1 并且小于或等于 N 的任一正整数, N 等于所述第一标识序列中的标识的个数。

在一种可能的设计中, 该字典密钥确定模块 301 还用于:

确定所述标识字典中的第 A 个子标识字典, 所述第 A 个子标识字典为根据所述脱敏密

钥中的第 a 个标识确定的子标识字典, $a = \begin{cases} m, B = n * (N * m) \\ [B/N] \% m, B \neq a * (N * m) \end{cases}$, B 为所述第一数

据在待脱敏的数据整体中的排列顺序, 所述数据整体包括至少两个数据, a 和 n 均为正整数, N 等于所述第一标识序列中的标识的个数;

根据所述第 A 个子标识字典确定第 A 标识序列中的第三标识, 所述第三标识在所述第 A 标识序列中位于所述第 a 个标识的后一位;

在所述脱敏密钥中将所述第 a 个标识替换为所述第三标识;

将替换得到的脱敏密钥确定为所述第一数据的下一个数据对应的脱敏密钥。

在一种可能的设计中, 所述脱敏密钥为与所述第一数据对应的脱敏任务相关的任意一个账号、任意一个号码、任意一个标识或者任意一个编号。

需要说明的是, 图 4 对应的实施例中未提及的内容可参见方法实施例的描述, 这里不再赘述。

本申请实施例中, 数据脱敏装置根据数据脱敏密钥确定标识字典, 然后利用数据字典

和标识字典对数据进行多次映射得到脱敏后的数据，在标识字典中对标识的映射相当于打乱了数据字典原始的映射顺序，使得映射具备随机映射的特性，其他人根据脱敏后的数据无法推导出映射关系，保证了数据的安全性，由于是通过数据脱敏密钥中的标识得到标识字典，那么根据脱敏密钥可确定标识字典，然后根据逆向的映射过程，可以对脱敏后的数据进行还原，保证了脱敏后的数据的可还原性。

参见图 5，图 5 是本申请实施例提供的另一种数据脱敏装置的组成结构示意图，该装置 40 包括处理器 401、存储器 402 以及输入输出接口 403。处理器 401 连接到存储器 402 和输入输出接口 403，例如处理器 401 可以通过总线连接到存储器 402 和输入输出接口 403。

处理器 401 被配置为支持数据脱敏装置执行前述方法实施例中的数据脱敏方法中相应的功能。该处理器 401 可以是中央处理器（central processing unit，CPU），网络处理器（network processor，NP），硬件芯片或者其任意组合。上述硬件芯片可以是专用集成电路（application specific integrated circuit，ASIC），可编程逻辑器件（programmable logic device，PLD）或其组合。上述 PLD 可以是复杂可编程逻辑器件（complex programmable logic device，CPLD），现场可编程逻辑门阵列（field-programmable gate array，FPGA），通用阵列逻辑（generic array logic，GAL）或其任意组合。

存储器 402 用于存储程序代码等。存储器 402 可以包括易失性存储器（volatile memory，VM），例如随机存取存储器（random access memory，RAM）；存储器 402 也可以包括非易失性存储器（non-volatile memory，NVM），例如只读存储器（read-only memory，ROM），快闪存储器（flash memory），硬盘（hard disk drive，HDD）或固态硬盘（solid-state drive，SSD）；存储器 402 还可以包括上述种类的存储器的组合。本申请实施例中，存储器 402 用于待脱敏的数据、数据字典、脱敏密钥等。

所述输入输出接口 403 用于输入或输出数据。

处理器 401 可以调用所述程序代码以执行以下操作：

确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个标识对应所述第一标识序列中的一个数据，所述数据脱敏密钥包括至少一个所述第一标识序列中的标识；

根据所述数据字典在所述第一标识序列中确定与所述第一数据对应的第一标识；

根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

根据所述数据字典确定在所述数据序列中与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

需要说明的是，各个操作的实现还可以对应参照方法实施例的相应描述；所述处理器 401 还可以与输入输出接口 403 配合执行上述方法实施例中的其他操作。

本申请实施例还提供一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被计算机执行时使所述计算机执行如前述实施例所述的方法，所述计算机可以为上述提到的数据脱敏装置的一部分。例如为上述的处理器 401。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，所述的存储介质可为磁碟、光盘、ROM 或 RAM 等。

以上所揭露的仅为本申请较佳实施例而已，当然不能以此来限定本申请之权利范围，因此依本申请权利要求所作的等同变化，仍属本申请所涵盖的范围。

权利要求

1、一种数据脱敏方法，其特征在于，包括：

确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个数据对应所述第一标识序列中的一个标识，所述脱敏密钥包括至少一个所述第一标识序列中的标识；

根据所述数据字典在所述第一标识序列中确定与所述第一数据对应的第一标识；

根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为根据所述脱敏密钥中的标识对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

根据所述数据字典在所述数据序列中确定与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

2、根据权利要求1所述的方法，其特征在于，所述根据所述脱敏密钥确定标识字典包括：

确定所述脱敏密钥中的各个标识对应的子标识字典；

按照预设的排列组合规则，分别将所述各个标识各自对应的子标识字典进行可重复性地排列和组合。

3、根据权利要求1~2任一所述的方法，其特征在于，所述脱敏密钥包括 m 个标识，所述标识字典包括 M 个子标识字典，所述 M 个子标识字典由可重复排列的 m 个子标识字典排列组合得到，所述 m 个子标识字典为根据所述 m 个标识确定的子标识字典， m 、 M 均为正整数， M 大于或等于 m ；

所述 M 个子标识字典中的第 i 个子标识字典包括所述第一标识序列和第 i 标识序列的对应关系，所述第 i 标识序列为对所述第一标识序列进行重排序得到的序列，所述第 i 标识序列中的一个标识对应所述第一标识序列中的一个标识，第 M 标识序列为所述第二标识序列， i 为正整数， i 大于或等于 1 并且小于或等于 M ；

所述根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识，包括：

根据所述第 i 个子标识字典在所述第 i 标识序列中确定与第 i 标识对应的第 $(i+1)$ 标识，其中，当 $i=1$ 时，所述第 i 标识为所述第一标识；

将第 M 标识序列中的第 $(M+1)$ 标识确定为所述第二标识。

4、根据权利要求3所述的方法，其特征在于，

$M=m$ ，所述第 i 个子标识字典为根据所述脱敏密钥中的第 j 个标识确定的子标识字典， $j=i$ ， j 为正整数， j 大于或等于 1 并且小于或等于 m 。

5、根据权利要求3所述的方法，其特征在于：

$M=2m-1$ ，所述第 i 个子标识字典为根据所述脱敏密钥中的第 j 个标识确定的子标识字典

$$\text{典, } j = \begin{cases} i, i \leq m \\ (i+1)/2, i > m \end{cases}.$$

6、根据权利要求4或5所述的方法，其特征在于，所述第*i*标识序列为对所述第一标识序列进行按顺序移位得到的序列，并且，所述第*j*个标识在所述第*i*个子标识字典的第*i*标识序列中位于第*k*位，*k*为大于或等于1并且小于或等于*N*的任一正整数，*N*等于所述第一标识序列中的标识的个数。

7、根据权利要求2-6任一项所述的方法，其特征在于，所述确定第一数据对应的数据字典和脱敏密钥之后，包括：

确定所述标识字典中的第*A*个子标识字典，所述第*A*个子标识字典为根据所述脱敏密

$$\text{钥中的第 } a \text{ 个标识确定的子标识字典, } a = \begin{cases} m, B = n * (N * m) \\ [B/N] \% m, B \neq a * (N * m) \end{cases}, B \text{ 为所述第一数}$$

据在待脱敏的数据整体中的排列顺序，所述数据整体包括至少两个数据，*a*和*n*均为正整数，*N*等于所述第一标识序列中的标识的个数；

根据所述第*A*个子标识字典确定第*A*标识序列中的第三标识，所述第三标识在所述第*A*标识序列中位于所述第*a*个标识的后一位；

在所述脱敏密钥中将所述第*a*个标识替换为所述第三标识；

将替换得到的脱敏密钥确定为所述第一数据的下一个数据对应的脱敏密钥。

8、根据权利要求1~7任一所述的方法，其特征在于，所述脱敏密钥为与所述第一数据对应的脱敏任务相关的任意一个账号、任意一个号码、任意一个标识或者任意一个编号。

9、一种数据脱敏装置，其特征在于，包括：

字典密钥确定模块，用于确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个数据对应所述第一标识序列中的一个标识，所述脱敏密钥包括至少一个所述第一标识序列中的标识；

第一标识确定模块，用于根据所述数据字典确定在所述第一标识序列中确定与所述第一数据对应的第一标识；

标识字典确定模块，用于根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为根据所述脱敏密钥中的标识对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

第二标识确定模块，用于根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

脱敏数据确定模块，用于根据所述数据字典在所述数据序列中确定与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

10、根据权利要求9所述的装置，其特征在于，所述标识字典确定模块，具体用于：

确定所述脱敏密钥中的各个标识对应的子标识字典；

按照预设的排列组合规则，分别将所述各个标识各自对应的子标识字典进行可重复性

地排列和组合。

11、根据权利要求 9~11 任一所述的装置，其特征在于，所述脱敏密钥包括 m 个标识，所述标识字典包括 M 个子标识字典，所述 M 个子标识字典由可重复排列的 m 个子标识字典排列组合得到，所述 m 个子标识字典为根据所述 m 个标识字典确定的子标识字典， m 、 M 均为正整数， M 大于或等于 m ；

所述 M 个子标识字典中的第 i 个子标识字典包括所述第一标识序列和第 i 标识序列的对应关系，所述第 i 标识序列为对所述第一标识序列进行重排序得到的序列，所述第 i 标识序列中的一个标识对应所述第一标识序列中的一个标识，第 M 标识序列为所述第二标识序列， i 为正整数， i 大于或等于 1 并且小于或等于 M ；

所述第二标识确定模块具体用于：

根据所述第 i 个子标识字典在所述第 i 标识序列中确定与所述第 i 标识对应的第 $(i+1)$ 标识，其中，当 $i=1$ 时，所述第 i 标识为所述第一标识；

将第 M 标识序列中的第 $(M+1)$ 标识确定为所述第二标识。

12、根据权利要求 11 所述的装置，其特征在于，

$M=m$ ，所述第 i 个子标识字典为根据所述脱敏密钥中的第 j 个标识确定的子标识字典， $j=i$ ， j 为正整数， j 大于或等于 1 并且小于或等于 m 。

13、根据权利要求 11 所述的装置，其特征在于，

$M=2m-1$ ，所述第 i 个子标识字典为根据所述脱敏密钥中的第 j 个标识确定的子标识字典， $j = \begin{cases} i, i \leq m \\ (i+1)/2, i > m \end{cases}$ 。

14、根据权利要求 12 或 13 所述的装置，其特征在于，所述第 i 标识序列为对所述第一标识序列进行按顺序移位得到的序列，并且，所述第 j 个标识在所述第 i 个子标识字典的第 i 标识序列中位于第 k 位， k 为大于或等于 1 并且小于或等于 N 的任一正整数， N 等于所述第一标识序列中的标识的个数。

15、根据权利要求 10~14 任一项所述的方法，其特征在于，所述字典密钥确定模块，还用于：

确定所述标识字典中的第 A 个子标识字典，所述第 A 个子标识字典为根据所述脱敏密

钥中的第 a 个标识确定的子标识字典， $a = \begin{cases} m, B = n * (N * m) \\ [B/N] \% m, B \neq a * (N * m) \end{cases}$ ， B 为所述第一数

据在待脱敏的数据整体中的排列顺序，所述数据整体包括至少两个数据， a 和 n 均为正整数， N 等于所述第一标识序列中的标识的个数；

根据所述第 A 个子标识字典确定第 A 标识序列中的第三标识，所述第三标识在所述第 A 标识序列中位于所述第 a 个标识的后一位；

在所述脱敏密钥中将所述第 a 个标识替换为所述第三标识；

将替换得到的脱敏密钥确定为所述第一数据的下一个数据对应的脱敏密钥。

16、根据权利要求 9~15 任一所述的装置，其特征在于，所述脱敏密钥为与所述第一数据对应的脱敏任务相关的任意一个账号、任意一个号码、任意一个标识或者任意一个编号。

17、一种数据脱敏装置，其特征在于，包括处理器、存储器以及通信接口，所述处理器、存储器和通信接口相互连接，其中，所述存储器用于存储计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令执行：

确定第一数据对应的数据字典和脱敏密钥，所述第一数据为待脱敏的数据，所述数据字典包括数据序列和第一标识序列的对应关系，所述数据序列包括多个数据，所述第一标识序列包括多个标识，所述数据序列中的一个数据对应所述第一标识序列中的一个标识，所述脱敏密钥包括至少一个所述第一标识序列中的标识；

根据所述数据字典在所述第一标识序列中确定与所述第一数据对应的第一标识；

根据所述脱敏密钥确定标识字典，所述标识字典包括所述第一标识序列和第二标识序列的对应关系，所述第二标识序列为根据所述脱敏密钥中的标识对所述第一标识序列进行重排序得到的序列，所述第二标识序列中的一个标识对应所述第一标识序列中的一个标识；

根据所述标识字典在所述第二标识序列中确定与所述第一标识对应的第二标识；

根据所述数据字典在所述数据序列中确定与所述第二标识对应的第二数据，并将所述第二数据确定为所述第一数据对应的脱敏数据。

18、根据权利要求 17 所述的装置，其特征在于，所述脱敏密钥包括 m 个标识，所述标识字典包括 M 个子标识字典，所述 M 个子标识字典由可重复排列的 m 个子标识字典排列组合得到，所述 m 个子标识字典为根据所述 m 个标识确定的子标识字典， m 、 M 均为正整数， M 大于或等于 m ；

所述 M 个子标识字典中的第 i 个子标识字典包括所述第一标识序列和第 i 标识序列的对应关系，所述第 i 标识序列为对所述第一标识序列进行重排序得到的序列，所述第 i 标识序列中的一个标识对应所述第一标识序列中的一个标识，第 M 标识序列为所述第二标识序列， i 为正整数， i 大于或等于 1 并且小于或等于 M ；

所述处理器具体用于：

根据所述第 i 个子标识字典在所述第 i 标识序列中确定与第 i 标识对应的第 $(i+1)$ 标识，其中，当 $i=1$ 时，所述第 i 标识为所述第一标识；

将第 M 标识序列中的第 $(M+1)$ 标识确定为所述第二标识。

19、根据权利要求 17 所述的装置，其特征在于，所述处理器还用于：

确定所述标识字典中的第 A 个子标识字典，所述第 A 个子标识字典为根据所述脱敏密

钥中的第 a 个标识确定的子标识字典， $a = \begin{cases} m, B = n * (N * m) \\ [B/N] \% m, B \neq a * (N * m) \end{cases}$ ， B 为所述第一数

据在待脱敏的数据整体中的排列顺序，所述数据整体包括至少两个数据， a 和 n 均为正整数， N 等于所述第一标识序列中的标识的个数；

根据所述第 A 个子标识字典确定第 A 标识序列中的第三标识，所述第三标识在所述第 A 标识序列中位于所述第 a 个标识的后一位；

在所述脱敏密钥中将所述第 a 个标识替换为所述第三标识；

将替换得到的脱敏密钥确定为所述第一数据的下一个数据对应的脱敏密钥。

20、一种计算机非易失性可读存储介质，其特征在于，所述计算机非易失性可读存储

介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如权利要求 1-8 任一项所述的方法。

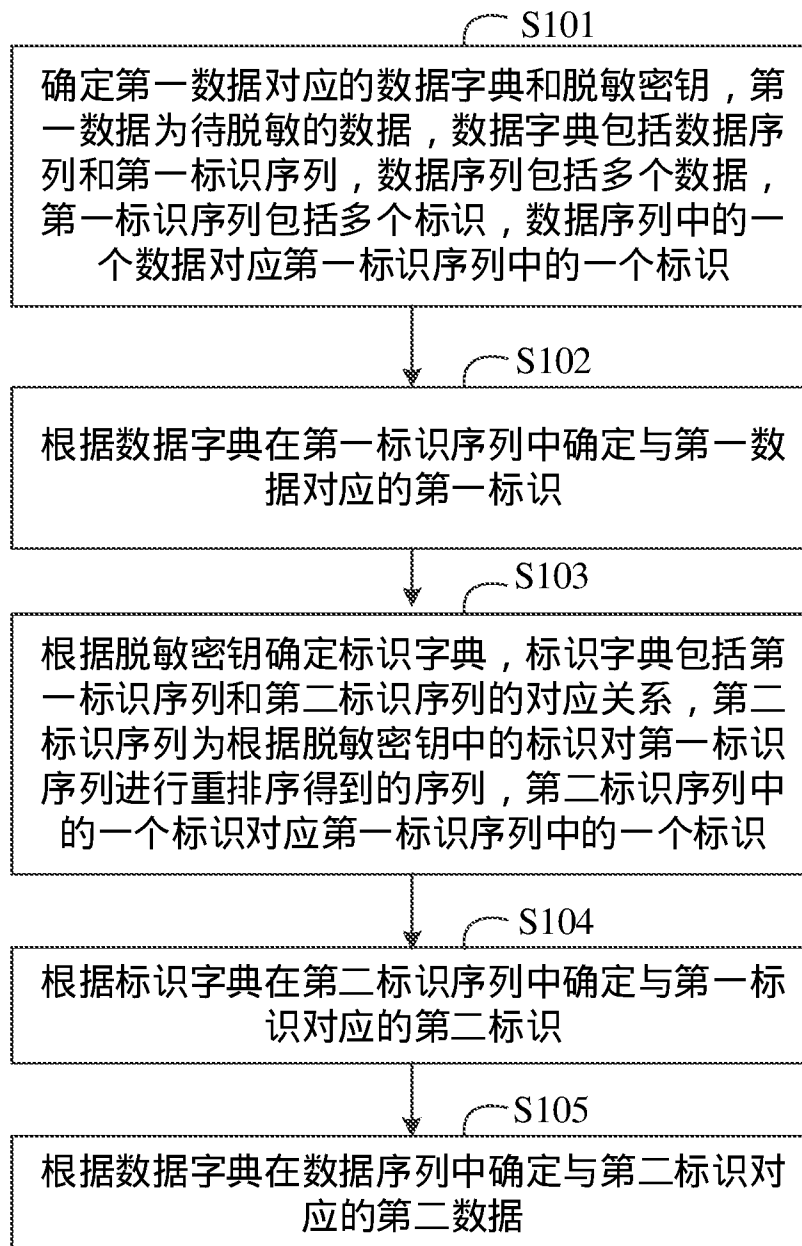


图 1

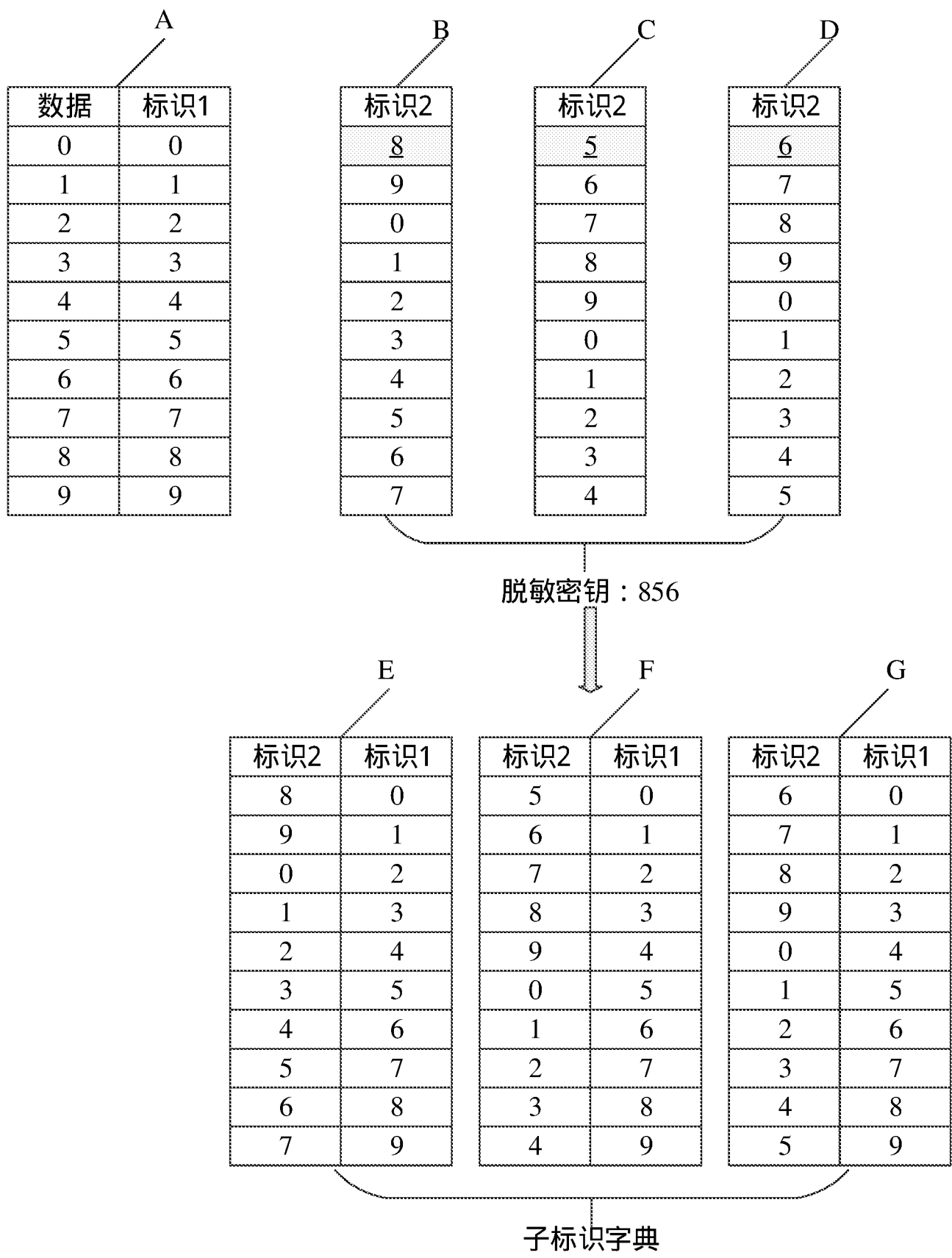


图 2

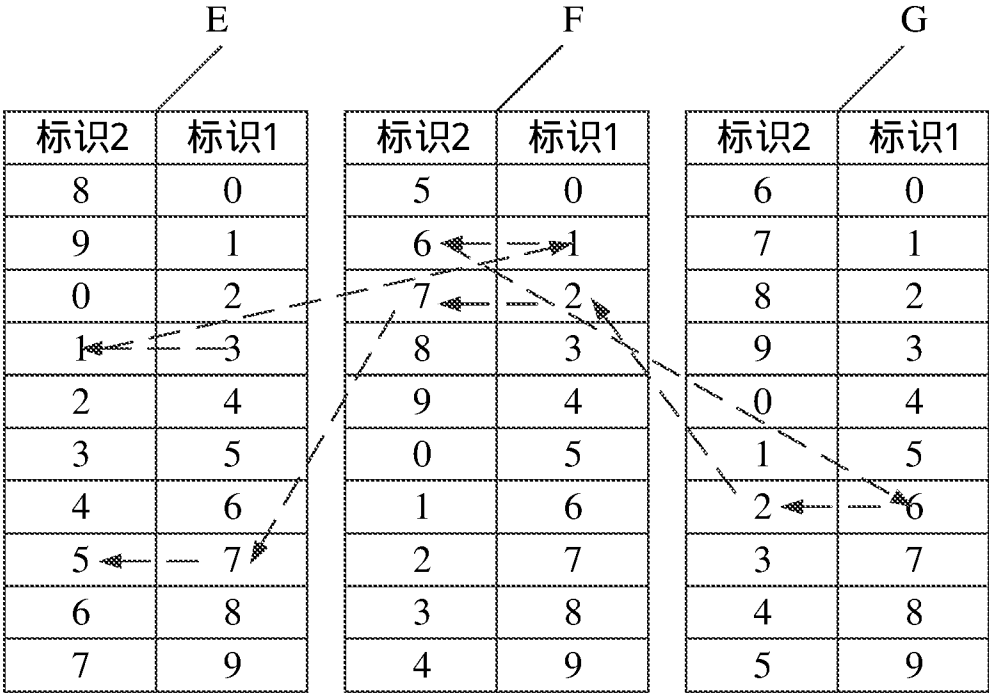


图 3

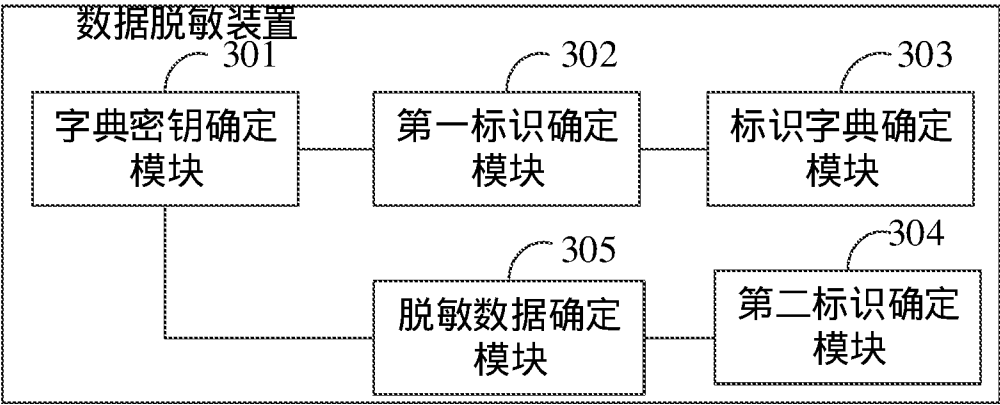


图 4

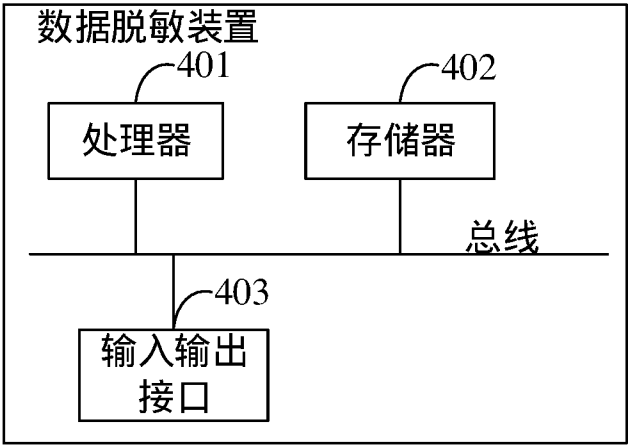


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102746

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/60(2013.01)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; H04L

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; CNKI; CNPAT: 数据, 脱敏, 字符, 数字, 序列, 字典, 排序, data, desensit+, char, character, number, sequence, dictionary, order

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110008744 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 12 July 2019 (2019-07-12) claims 1-10, description, paragraphs [0038], [0061]	1-20
A	CN 109039586 A (GUIZHOU UNIVERSITY) 18 December 2018 (2018-12-18) description, paragraphs [0020]-[0028]	1-20
A	CN 107609418 A (SHENZHEN NIUDINGFENG TECHNOLOGY CO., LTD.) 19 January 2018 (2018-01-19) entire document	1-20
A	US 2010228783 A1 (CASTELLANOS, Maria G. et al.) 09 September 2010 (2010-09-09) entire document	1-20



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

03 December 2019

Date of mailing of the international search report

02 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/102746

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	110008744	A	12 July 2019	None	
CN	109039586	A	18 December 2018	None	
CN	107609418	A	19 January 2018	None	
US	2010228783	A1	09 September 2010	None	

国际检索报告

国际申请号

PCT/CN2019/102746

A. 主题的分类

G06F 21/60 (2013.01) 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; H04L

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; EPDOC; CNKI; CNPAT: 数据, 脱敏, 字符, 数字, 序列, 字典, 排序, data, desensit+, char, character, number, sequence, dictionary, order

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 110008744 A (平安科技深圳有限公司) 2019年 7月 12日 (2019 - 07 - 12) 权利要求1-10, 说明书第[0038], [0061]段	1-20
A	CN 109039586 A (贵州大学) 2018年 12月 18日 (2018 - 12 - 18) 说明书第[0020]-[0028]段	1-20
A	CN 107609418 A (深圳市牛鼎丰科技有限公司) 2018年 1月 19日 (2018 - 01 - 19) 全文	1-20
A	US 2010228783 A1 (CASTELLANOS, Maria G. 等) 2010年 9月 9日 (2010 - 09 - 09) 全文	1-20

☐ 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

2019年 12月 3日

国际检索报告邮寄日期

2020年 1月 2日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (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

授权官员

庄湧

电话号码 86-(10)-53961296

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/102746

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	110008744	A	2019年 7月 12日	无	
CN	109039586	A	2018年 12月 18日	无	
CN	107609418	A	2018年 1月 19日	无	
US	2010228783	A1	2010年 9月 9日	无	