(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 8 月 11 日 (11.08.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/124070 A1

(51) 国际专利分类号:
H03M 7/30 (2006.01)

(21) 国际申请号: PCT/CN2016/070805

(22) 国际申请日: 2016 年 1 月 13 日 (13.01.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510059809.6 2015 年 2 月 4 日 (04.02.2015) CN

(71) 申请人: **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: **朱金伟 (ZHU, Jinwei)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **严龙 (YAN, Long)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **周庆庆 (ZHOU, Qingqing)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: **北京龙双利达知识产权代理有限公司 (LONGSUN LEAD IP LTD.)**; 中国北京市海淀区丹棱街 16 号海兴大厦 C 座 1108, Beijing 100080 (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)。

[见续页]

(54) **Title**: DATA PROCESSING METHOD AND DEVICE

(54) **发明名称** : 一种数据处理的方法和设备

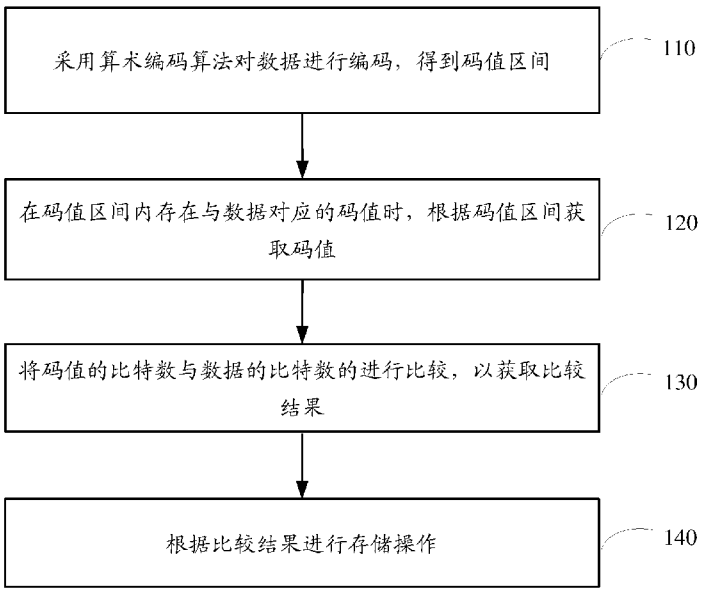


图 1

110 Code data by using an arithmetic coding algorithm to obtain a code value range
120 When a code value corresponding to the data exists in the code value range, acquire the code value according to the code value range
130 Compare a bit number of the code value with a bit number of the data to obtain a comparison result
140 Conduct a storing operation according to the comparison result

(57) **Abstract**: Provided in an embodiment of the present invention are a data processing method and device, the method comprising: coding data by using an arithmetic coding algorithm to obtain a code value range; when a code value corresponding to the data exists in the code value range, acquiring the code value according to the code value range; comparing a bit number of the code value with a bit number of the data to obtain a comparison result; and conducting a storing operation according to the comparison result. The embodiment of the present invention reduces data storage space.

(57) **摘要**: 本发明实施例提供了一种数据处理的方法和设备，该方法包括采用算术编码算法对数据进行编码，得到码值区间；在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；根据比较结果进行存储操作。本发明实施例能够减小数据的存储空间。

WO 2016/124070 A1

根据细则 4.17 的声明:

— 关于申请人有权申请并被授予专利(细则 4.17(ii))

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种数据处理的方法和设备

技术领域

本发明涉及数据处理领域，特别涉及一种数据处理的方法和设备。

背景技术

算术编码是近年来另一种非常实用的无损压缩算法。算法编码的核心思想在于将源数据中所有可能出现的被编码符号映射到一个整数集合，并赋予每个编码符号相应的出现概率(在此要求所有字符出现的概率和为 1)。依照每个字符的出现概率，各字符占据[0,1]区间中的一个半开半闭的连续区间，区间长度值即概率值，同时区间之间彼此相互独立。接着按照映射表将所需编码字符串映射成一个整数序列，然后根据所需要编码的源数据中被编码符号出现的概率，逐步将源数据转变成一个对应于[0,1]区间的实数区间，并取该区间中的一个实数作为码值存储在计算机中。下次进行编码的区间是上次编码所得到的区间，而每次所有符号出现的概率比值不变。在解码的时候将该二进制码值按照逆变化还原成相应的整数序列，然后再映射回原来的字符串。例如，对于整数集空间{0,1,2,3}，出现概率分布为{0.2, 0.5, 0.2, 0.1}。那么对应输入序列为<210013>的数据，编码区间依次如下:[0.7，0.9]、[0.74，0.84]、[0.74，0.76]、[0.74，0.744]、[0.7408，0.7428]、[0.7426，0.7428]，最终该数据对应的码值区间为[0.7426，0.7428]（最后一个字符序列对应的编码区间），该数据的码值为[0.7426，0.7428]中的一个数值。

对于待编码数据，现有算术编码中不考虑压缩是否具有收益，直接对数据进行压缩，然后存储对应的算术编码后的码值，由于某些数据对应的码值比特数较大，因此，现有技术会增大数据的存储空间。

发明内容

本发明实施例提供一种数据处理的方法和设备，该方法能够减小数据的存储空间。

第一方面，提供了一种数据处理的方法，包括：采用算术编码算法对数据进行编码，得到码值区间；在该码值区间内存在与该数据对应的码值时，根据该码值区间获取该码值；将该码值的比特数与该数据的比特数的进行比

较，以获取比较结果；根据该比较结果进行存储操作。

结合第一方面，在第一种可能的实现方式中，该比较结果为该码值的比特数小于该数据的比特数，其中，该根据该比较结果进行存储操作，包括：根据该比较结果，存储该码值。

结合第一方面，在第二种可能的实现方式中，该比较结果为该码值的比特数大于或等于该数据的比特数，其中，该根据该比较结果进行存储操作，包括：根据该比较结果，存储该数据。

结合第一种可能的实现方式，在第三种可能的实现方式中，还包括：根据该码值进行该数据的应用操作，该应用操作包括等值比较、排列和模糊查询中的至少一种。

结合第三种可能的实现方式，在第四种可能的实现方式中，该数据为标识 ID 类字符串，该应用操作包括等值比较，该根据该码值进行该数据的应用操作，包括：在该码值与待比较的码值相等时，确定该数据与该待比较的码值对应的数据为相同数据。

结合第三种可能的实现方式，在第五种可能的实现方式中，该数据为 ID 类字符串或为字段的字母字符串，该应用操作包括排列，该根据该码值进行该数据的应用操作，包括：根据该码值的大小，确定该码值在待排列码值中的位置，该码值的位置用于表示该数据在与该待排列的码值对应的数据中的位置。

结合第三种可能的实现方式，在第六种可能的实现方式中，该应用操作包括模糊查询，该根据该码值进行该数据的应用操作，包括：根据该码值是否在所需模糊查询的前缀字符串的编码区间中，确定该数据是否包括该前缀字符串，其中，在该码值在所需模糊查询的前缀字符串的编码区间中时，该数据包括该前缀字符串，在该码值不在所需模糊查询的前缀字符串的编码区间中时，该数据不包括该前缀字符串。

结合第一方面、第一至第六种可能的实现方式中的任一种可能的实现方式，在第七种可能的实现方式中，该采用算术编码算法对数据进行编码，得到码值区间，包括：采用算术编码算法对该数据进行编码，得到编码区间；对该数据的编码区间进行重扩展，得到重扩展后的编码区间；根据该重扩展后的编码区间，采用算术编码算法继续对该数据进行编码，得到码值区间。

结合第七种可能的实现方式，在第八种可能的实现方式中，该对该数据

的编码区间进行重扩展，得到重扩展后的编码区间，包括：当满足以下条件中的至少一个时，对该数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的该数据的字符个数达到字符个数阈值，和该数据的编码区间的长度小于区间阈值。

结合第一方面、第一至第八种可能的实现方式中的任一种可能的实现方式，在第九种可能的实现方式中，在该码值区间内存在与该数据对应的码值时，根据该码值区间获取该码值之前，还包括：确定该码值区间内是否存在与该数据对应的合适的码值。

结合第九种可能的实现方式，在第十种可能的实现方式中，还包括：在该码值区间内不存在与该数据对应的合适的码值时，存储该数据。

第二方面，提供了一种数据处理的方法，包括：采用算术编码算法对数据进行编码，得到编码区间；对该数据的编码区间进行重扩展，得到重扩展后的编码区间；根据该重扩展后的编码区间，采用算术编码算法继续对该数据进行编码，得到码值区间；根据该码值区间获取该码值；存储该码值。

结合第二方面，在第一种可能的实现方式中，该对该数据的编码区间进行重扩展，得到重扩展后的编码区间，包括：当满足以下条件中的至少一个时，对该数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的该数据的字符个数达到字符个数阈值，和该数据的编码区间的长度小于区间阈值。

第三方面，提供了一种数据处理的方法，包括：获取数据的码值和重扩展字符位置；采用算术编码算法对数据的码值进行解码，得到解码区间；根据该重扩展字符位置，对该数据的解码区间进行重缩放，得到重缩放的解码区间；根据该重缩放的解码区间，采用算术编码算法继续对该数据进行解码，得到该数据。

结合第三方面，在第一种可能的实现方式中，该根据重扩展字符位置，对该数据的解码区间进行重缩放，得到重缩放的解码区间，包括：根据重扩展字符位置，确定重缩放字符位置，其中该重扩展字符位置与该重缩放字符位置相倒置；根据重缩放字符位置对该数据的解码区间进行重缩放，得到重缩放的解码区间。

第四方面，提供了一种数据处理的方法，包括：采用算术编码算法对数据进行编码，得到码值区间；根据该码值区间获取该码值；存储该码值；根

据该码值进行该数据的应用操作，该应用操作包括等值比较、排列和模糊查询中的至少一种。

结合第四方面，在第一种可能的实现方式中，该数据为 ID 类字符串，该应用操作包括等值比较，该根据该码值进行该数据的应用操作，包括在该码值与待比较的码值相等时，确定该数据与该待比较的码值对应的数据为相同数据。

结合第四方面，在第二种可能的实现方式中，该数据为 ID 类字符串或为字段的字母字符串，该应用操作包括排列，该根据该码值进行该数据的应用操作，包括：根据该码值的大小，确定该码值在待排列码值中的位置，该码值的位置用于表示该数据在与该待排列的码值对应的数据中的位置。

结合第四方面，在第三种可能的实现方式中，该应用操作包括模，该根据该码值进行该数据的应用操作，包括：根据该码值是否在所需模糊查询的前缀字符串的编码区间中，确定该数据是否包括该前缀字符串，其中，在该码值在所需模糊查询的前缀字符串的编码区间中时，该数据包括该前缀字符串在该码值不在所需模糊查询的前缀字符串的编码区间中时，该数据不包括该前缀字符串。

第五方面，提供了一种数据处理的设备，包括：编码单元，用于采用算术编码算法对数据进行编码，得到码值区间；获取单元，用于在该码值区间内存在与该数据对应的码值时，根据该码值区间获取该码值；比较单元，用于将该码值的比特数与该数据的比特数的进行比较，以获取比较结果；第一存储单元，用于根据该比较结果进行存储操作。

结合第五方面，在第一种可能的实现方式中，该比较结果为该码值的比特数小于该数据的比特数，其中，该第一存储单元根据该比较结果，存储该码值。

结合第五方面，在第二种可能的实现方式中，该比较结果为该码值的比特数大于或等于该数据的比特数，其中，该第一存储单元根据该比较结果，存储该数据。

结合第五方面的第一种可能的实现方式，在第三种可能的实现方式中，还包括：应用单元，用于根据该码值进行该数据的应用操作，该应用操作包括等值比较、排列和模糊查询中的至少一种。

结合第五方面的第三种可能的实现方式，在第四种可能的实现方式中，

该数据为 ID 类字符串，该应用操作包括等值比较，该应用单元在该码值与待比较的码值相等时，确定该数据与该待比较的码值对应的数据为相同数据。

结合第五方面的第三种可能的实现方式，在第六种可能的实现方式中，该数据为 ID 类字符串或为字段的字母字符串，该应用操作包括排列，该应用单元根据该码值的大小，确定该码值在待排列码值中的位置，该码值的位置用于表示该数据在与该待排列的码值对应的数据中的位置。

结合第五方面的第三种可能的实现方式，在第八种可能的实现方式中，该应用操作包括模糊查询，该应用单元根据该码值是否在所需模糊查询的前缀字符串的编码区间中确定该数据是否包括该前缀字符串，其中，在该码值在所需模糊查询的前缀字符串的编码区间中时，该数据包括该前缀字符串，在该码值不在所需模糊查询的前缀字符串的编码区间中时，该数据不包括该前缀字符串。

结合第五方面、第五方面的第一至第九种可能的实现方式中的任一种可能的实现方式，在第十种可能的实现方式中，该编码单元采用算术编码算法对该数据进行编码，得到编码区间；对该数据的编码区间进行重扩展，得到重扩展后的编码区间；根据该重扩展后的编码区间，采用算术编码算法继续对该数据进行编码，得到码值区间。

结合第五方面的第十种可能的实现方式，在第十一种可能的实现方式中，该编码单元当满足以下条件中的至少一个时，对该数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的该数据的字符个数达到字符个数阈值，和该数据的编码区间的长度小于区间阈值。

结合第五方面、第五方面的第一至第十一种可能的实现方式中的任一种可能的实现方式，在第十二种可能的实现方式中，还包括：确定单元，用于在该获取单元获取该码值之前，确定该码值区间内是否存在与该数据对应的合适的码值。

结合第五方面的第十二种可能的实现方式，在第十三种可能的实现方式中，还包括：第二存储单元，用于在该码值区间内不存在与该数据对应的合适的码值时，存储该数据。

第六方面，提供了一种数据处理的设备，包括：第一编码单元，用于采

用算术编码算法对数据进行编码，得到编码区间；扩展单元，用于对该数据的编码区间进行重扩展，得到重扩展后的编码区间；第二编码单元，用于根据该重扩展后的编码区间，采用算术编码算法继续对该数据进行编码，得到码值区间。获取单元，用于根据该码值区间获取该码值；存储单元，用于存储该码值。

结合第六方面，在第一种可能的实现方式中，该扩展单元当满足以下条件中的至少一个时，对该数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的该数据的字符个数达到字符个数阈值，和该数据的编码区间的长度小于区间阈值。

第七方面，提供了一种数据处理的设备，包括：第一获取单元，用于获取数据的码值和重扩展字符位置；第一解码单元，用于采用算术编码算法对数据的码值进行解码，得到解码区间；缩放单元，用于根据该重扩展字符位置，对该数据的解码区间进行重缩放，得到重缩放的解码区间；第二解码单元，用于根据该重缩放的解码区间，采用算术编码算法继续对该数据进行解码，得到该数据。

结合第七方面，在第一种可能的实现方式中，该缩放单元根据重扩展字符位置，确定重缩放字符位置，其中该重扩展字符位置与该重缩放字符位置相倒置；根据重缩放字符位置对该数据的解码区间进行重缩放，得到重缩放的解码区间。

第八方面，提供了一种数据处理的设备，包括：编码单元，用于采用算术编码算法对数据进行编码，得到码值区间；获取单元，用于根据该码值区间获取该码值；存储单元，用于存储该码值；应用单元，用于根据该码值进行该数据的应用操作，该应用操作包括等值比较、排列和模糊查询中的至少一种。

结合第八方面，在第一种可能的实现方式中，该数据为 ID 类字符串，该应用操作包括等值比较，该应用单元在该码值与待比较的码值相等时，确定该数据与该待比较的码值对应的数据为相同数据。

结合第八方面，在第三种可能的实现方式中，该数据为 ID 类字符串或为字段的字母字符串，该应用操作包括排列，该应用单元根据该码值的大小，确定该码值在待排列码值中的位置，该码值的位置用于表示该数据在与该待排列的码值对应的数据中的位置。

结合第八方面，在第五种可能的实现方式中，该应用操作包括模糊查询，该应用单元根据该码值是否在所需模糊查询的前缀字符串的编码区间中，确定该数据是否包括该前缀字符串，其中，在该码值在所需模糊查询的前缀字符串的编码区间中时，该数据包括该前缀字符串，在该码值不在所需模糊查询的前缀字符串的编码区间中时，该数据不包括该前缀字符串。

基于上述技术方案，本发明实施例通过采用算术编码算法对数据进行编码，得到码值区间。在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；根据比较结果进行存储操作。本发明实施例能够减小数据的存储空间。

附图说明

为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例中所需要使用的附图作简单地介绍，显而易见地，下面所描述的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 是根据本发明一个实施例的用于数据压缩的方法的示意性流程图。

图 2 是根据本发明一个实施例的字段排序示意图。

图 3 是根据本发明一个实施例的模糊查询示意图。

图 4 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 5 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 6 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 7 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 8 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 9 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 10 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 11 是根据本发明另一实施例的数据处理的方法的示意性流程图。

图 12 是根据本发明一个实施例的数据处理的设备的示意框图。

图 13 是根据本发明另一实施例的数据处理的设备的示意框图。

图 14 是根据本发明另一实施例的数据处理的设备的示意框图。

图 15 是根据本发明另一实施例的数据处理的设备的示意框图。

图 16 是根据本发明另一实施例的数据处理的设备的示意框图。

图 17 是根据本发明另一实施例的数据处理的设备的示意框图。

图 18 是根据本发明另一实施例的数据处理的设备的示意框图。

图 19 是根据本发明另一实施例的数据处理的设备的示意框图。

具体实施方式

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明的一部分实施例，而不是全部实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都应属于本发明保护的范围。

图 1 是根据本发明一个实施例的用于数据压缩的方法的示意性流程图。图 1 所示的方法可以由图 4 中的数据压缩的设备执行，如图 1 所示的方法包括:

110，采用算术编码算法对数据进行编码，得到码值区间。

具体地，码值区间可以为该数据的最后一个字符串对应的编码区间。

应理解，本发明实施例中的数据可以是标识（Identity，ID）类字符串或字段的字母字符串等，其中，ID 类字符串可以包括: 数据单号、图书索引号、公交车牌号、产品单据号、国际移动用户识别码（International Mobile Subscriber Identification Number，IMSI）或移动设备国际身份码（International Mobile Equipment Identity，IMEI）等; 字母字符串可以包括中文、韩文、日文对应的音标等。

120，在码值区间内存在与数据对应的码值时，根据码值区间获取码值。

换句话说，在码值区间内可以取到合适的码值时，根据码值区间获取码值。例如，合适的码值可以为满足一定位数的码值，例如，码值的二进制位不超过 16 位、32 位或 64 位等。

应注意，可以根据现有方法根据码值区间获取码值，也可以根据预设条件获取码值，例如，预设条件为码值的二进制位不超过 16 位、32 位或 64 位等，本发明实施例并不对此做限定。

130，将码值的比特数与数据的比特数的进行比较，以获取比较结果。

具体地，比较结果可以包括码值的比特数小于数据的比特数、码值的比特数等于数据的比特数或码值的比特数大于数据的比特数。

140，根据比较结果进行存储操作。

具体地，采用算术编码算法对数据进行编码，得到码值区间，然后，确定该码值区间（数据的最后一个字符串对应的编码区间）内是否可以取到合适的码值，如果无法取到将直接存储数据而无需算术编码。如果可以取到合适的码值，然后将码值的比特数与数据的比特数的进行比较（判断压缩收益），具体地，判断该合适的码值的比特数是否大于或等于表示原数据的比特数，如果该合适的码值的比特数大于或等于原数据的比特数，也就是无压缩收益，则放弃采用算术编码可以直接存储该数据；如果该合适的码值的比特数小于原数据的比特数，也就是有压缩收益，则存储码值。

因此，本发明实施例通过采用算术编码算法对数据进行编码，得到码值区间。在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；根据比较结果进行存储操作。本发明实施例能够降低错误的编码，减小数据的存储空间。

可选地，作为另一实施例，比较结果为码值的比特数小于数据的比特数，其中，在 140 中，根据比较结果，存储码值。

具体地，当码值的比特数小于数据的比特数时，表明存在压缩收益，存储码值。

可替代地，作为另一实施例，比较结果为码值的比特数大于或等于数据的比特数，其中，在 140 中，根据比较结果，存储数据。

具体地，当码值的比特数大于或等于数据的比特数时，表明无压缩收益，存储数据。

可选地，作为另一实施例，当存储码值时，本发明实施例方法还包括根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种。

具体地，采用算术编码算法对数据进行编码，得到码值区间；在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；其中比较结果为码值的比特数小于数据的比特数，存储码值，根据码值进行数据的等值比较、排列和模糊查询中的至少一种。例如，可以根据码值进行数据的等值比较、根据码值进行数据的排列，或者可以根据码值对数据进行模糊查询。下面分别介绍上述根据码值进行数据的应用操作。

具体地，作为另一实施例，数据为 ID 类字符串，应用操作包括等值比

较，根据码值进行数据的应用操作，包括：根据码值进行数据的等值比较。

进一步地，作为另一实施例，根据码值进行数据的等值比较，包括：在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。

应理解，待比较的码字为待比较的数据（与待比较的码值对应的数据）的码值。具体地，进行码值的等值比较可以理解为进行数据的匹配，例如，当两个数据的码值相等时，可以确定该两个码值对应的该两个数据为相同数据，也即匹配成功；当两个码值不相等时，可以确定该两个码值对应的该两个数据为不同数据，也即匹配不成功。

可替代地，作为另一实施例，数据为 ID 类字符串或为字段的字母字符串，应用操作包括排列，根据码值进行数据的应用操作，包括：根据码值进行数据的排列。

进一步地，作为另一实施例，根据码值进行数据的排序，包括：根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。

具体的，例如，进行数据的排序可以理解为对多个数据进行排序。例如，有 5 个数据，对应 5 个码值，5 个码值由小到大排序，例如，当前码值为 5 个码值中的第 4 个，那么第 4 个码值对应的数据排在该 5 个数据中的第 4 位。

在现有的数据库实现中，通常会将压缩技术与查询操作分开考虑，即单独考虑数据存储技术以及相关的查询优化技术。数据库的重要作用在于存储记录所关心事物的重要描述，事物发展等相关信息。因此当事物描述较为冗繁时不仅占用大量的存储空间，也会带来查询上的不便。现有对数据在进行多种查询操作时(如字符串比较，字符串排序等)，因需要逐步比较字符串中的字符，会导致查询效率较低。而本发明实施例无需算术解码的过程，直接根据数据对应的码值进行数据的比较(匹配)和/或根据码值进行数据的排列，将原本复杂数据类型的查询操作等效成码值的查询，快速简单。

例如，当数据中只出现数字或字母的应用场景中(此时概率将全分给数字或全分给字母)。例如，数据为非拉丁字母文字。如中文、韩文、日文都有对应的音标。可以通过将对应文字转换成其音标表示方式，即只包含字母的字符串，在采用算术编码算法对数据进行编码，得到码值区间；在码值区间内存在与数据对应的码值时，确定码值的比特数小于数据的比特数；存储

码值。根据码值对数据对应的字段进行排序。

例如，中文拼音字母出现概率分布(全拼)，A(0.107)，B(0.014)，C(0.017)，D(0.030)，E(0.062)，F(0.009)，G(0.060)，H(0.067)，I(0.141)，J(0.023)，K(0.008)，L(0.017)，M(0.014)，N(0.117)，O(0.065)，P(0.008)，Q(0.013)，R(0.006)，S(0.026)，T(0.015)，U(0.096)，V(0.001)，W(0.010)，X(0.020)，Y(0.028)，Z(0.026)。如图 2 所示，对于字段“优秀”“良好”“及格”，按字母顺序排列并编码，其中，“优秀”对应的字母字符串为“youxiu”，对应的码值为 0.96684845；“良好”对应的字母字符串为“lianghao”，对应的码值为 0.544375656；“及格”对应的字母字符串为“jige”，对应的码值为 0.516228。按照码值由小到大的依次排序为 0.516228、0.544375656 和 0.96684845，分别对应“及格”、“良好”和“优秀”。

可替代地，作为另一实施例，应用操作包括模糊查询，根据码值进行数据的应用操作，包括：根据码值对数据进行模糊查询。

进一步地，作为另一实施例，根据码值对数据进行模糊查询，包括：根据码值是否在所需模糊查询的前缀字符串的编码区间中，确定数据是否包括前缀字符串，其中，在码值在所需模糊查询的前缀字符串的编码区间中时，数据包括前缀字符串，在码值不在所需模糊查询的前缀字符串的编码区间中时，数据不包括前缀字符串。

换句话说，码值在所需模糊查询的前缀字符串的编码区间中时，数据满足模糊查询，码值不在所需模糊查询的前缀字符串的编码区间中时，数据不满足模糊查询。

具体而言，当字符串经过算术编码压缩后，得到的是一系列的码值。码值来自于对字符串编码后所得到的编码区间，而编码区间彼此之间相互不正交。我们还注意到字符串的编码区间总是包含于其前缀字符串的编码区间。例如，字符串‘A12986572’的编码区间一定包含于‘A1298’、‘A12’等前缀字符串的编码区间中。只需判断码值是否在所需模糊查询的前缀字符串的编码区间中，就可进行模糊查询。例如，整数集空间{0,1,2,3}，出现概率分布{0.2, 0.5, 0.2, 0.1}。针对模糊查询%210xxx 而言，对于数据 212132、210312、210231 和 211123 进行该模糊查询:如图 3 所示，“210”的码值区间为[0.74，0.76]，212132、210312、210231 和 211123 对应的码值区间分别为 0.8238、0.7592、0.7576 和 0.7923，由于 0.7592 和 0.7576 在模糊查询的前缀字符串的编码区

间[0.74，0.76]中，0.8238 和 0.7923 不在模糊查询的前缀字符串的编码区间[0.74，0.76]中，因此，210312 和 210231 满足模糊查询，212132 和 211123 不满足模糊查询。

因此，本发明实施例另外在模糊查询操作中，当索引字符数大于 2 时，即可抵消判断区间定位所需操作。

可选地，作为另一实施例，在 110 中，采用算术编码算法对数据进行编码，得到编码区间；对数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。其中，记录相应重扩展字符位置用于解码设备根据码值和重扩展字符位置得到重缩放的字符位置，并根据重缩放的字符位置进行解码区间的重缩放，最终得到数据。

具体地，本发明实施例可以对任一个字符对应的编码区间进行重扩展，也可以根据预设条件进行编码区间的重扩展。

因此，本发明实施例可以对数据的编码区间进行编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

进一步地，作为另一实施例，在 110 中，当满足以下条件中的至少一个时，对数据的编码区间进行重扩展，得到重扩展后的编码区间：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。

具体而言，当满足数据的字符个数大于预设字符个数阈值时，在 110 中，采用算术编码算法对数据进行编码，当已完成编码的数据的字符个数达到预设字符个数阈值时，对与预设字符个数阈值对应的数据的字符的编码区间进行重扩展，并记录相应重扩展字符位置；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

换句话说，采用算术编码算法对数据的字符串逐步进行编码时，当已完成编码的数据的字符个数达到预设字符个数阈值时，对与预设字符个数阈值对应的数据的字符的编码区间进行重扩展，并记录相应重扩展字符位置；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值

区间。

例如，当前数据包括 12 个字符串，预设字符个数阈值为 10，那么在采用算术编码算法对当前数据的 12 个字符串进行编码时，对 12 个字符串中的第 10 个字符的编码区间重扩展，并记录重扩展的字符位置为第 10 个字符的位置，然后，重扩展后的编码区间，采用算术编码算法继续对数据（11 和 12 个字符）进行编码，最终得到码值区间（第 12 字符对应的编码区间）。

随着数据的字符串长度的增加，编码所得到的编码区间长度也会越来越小，在较小的编码区间长度中不容易取到合适的码值。因此，为了尽可能避免可能出现的无法正确进行算术编码的问题，本发明实施例在字符个数达到预设字符个数阈值时进行编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

当满足数据的编码区间的长度小于预设的阈值时，在 110 中，对数据的编码区间进行重扩展；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

换句话说，采用算术编码算法对数据的字符串逐步进行编码时，编码区间会越来越小，当编码区间小于预设阈值时，对小于预设阈值的编码区间进行重扩展；之后，根据重扩展后的编码区间，采用算术编码算法继续对数据中为完成的字符串进行编码，最后得到码值区间。

例如，当前数据包括 12 个字符串，预设阈值为 0.05，那么在采用算术编码算法对当前数据的 12 个字符串进行编码时，当编码区间小于 0.05 时，对数据的编码区间进行重扩展；例如，当第 7 个字符的编码区间为 0.1，第 8 个字符的编码区间为 0.04，那么，对 8 个字符的编码区间 0.04 进行重扩展，例如，将 0.04 扩展为 1 或 10 等，之后，根据重扩展后的编码区间（1 或 10）采用算术编码算法继续对数据（第 9 至 12 字符）进行编码，得到码值区间（第 12 字符对应的编码区间）。

随着数据的字符串长度的增加，编码所得到的编码区间长度也会越来越小，在较小的编码区间长度中不容易取到合适的码值。因此，为了尽可能避免可能出现的无法正确进行算术编码的问题，本发明实施例通过对编码区间

的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

应注意，进行区间扩展的所有信息将同样发送给解码器，即编码器将二进制码值传送给解码器的同时，也会将移位机制信息发送给解码器。用于同步信息，保证在解码的时候得到正确的解码结果。

图 4 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 4 所示的方法可以由数据处理的设备执行，具体地可以由编码设备执行，如图 4 所示，该方法包括:

410，采用算术编码算法对数据进行编码，得到编码区间;

420，对数据的编码区间进行重扩展，得到重扩展后的编码区间;

具体地，本发明实施例可以对任一个字符对应的编码区间进行重扩展，也可以根据预设条件进行编码区间的重扩展。

430，根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

440，根据码值区间获取码值。

450，存储码值。

因此，本发明实施例通过对数据的编码区间进行编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

进一步地，作为另一实施例，在 420 中，当满足以下条件中的至少一个时，对数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。其中，记录相应重扩展字符位置用于解码设备根据码值和重扩展字符位置得到重缩放的字符位置，并根据重缩放的字符位置进行解码区间的重缩放，最终得到数据。

具体而言，当满足数据的字符个数大于预设字符个数阈值时，采用算术编码算法对数据进行编码，当已完成编码的数据的字符个数达到预设字符个

数阈值时，对与预设字符个数阈值对应的数据的字符的编码区间进行重扩展；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

换句话说，采用算术编码算法对数据的字符串逐步进行编码时，当已完成编码的数据的字符个数达到预设字符个数阈值时，对与预设字符个数阈值对应的数据的字符的编码区间进行重扩展；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

例如，当前数据包括 12 个字符串，预设字符个数阈值为 10，那么在采用算术编码算法对当前数据的 12 个字符串进行编码时，对 12 个字符串中的第 10 个字符的编码区间重扩展，并记录重扩展的字符位置为第 10 个字符的位置，然后，重扩展后的编码区间，采用算术编码算法继续对数据（11 和 12 个字符）进行编码，最终得到码值区间（第 12 字符对应的编码区间）。

随着数据的字符串长度的增加，编码所得到的编码区间长度也会越来越小，在较小的编码区间长度中不容易取到合适的码值。因此，为了尽可能避免可能出现的无法正确进行算术编码的问题，本发明实施例在字符个数达到预设字符个数阈值时进行编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

当满足数据的编码区间的长度小于预设的阈值时，对数据的编码区间进行重扩展；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

换句话说，采用算术编码算法对数据的字符串逐步进行编码时，编码区间会越来越小，当编码区间小于预设阈值时，对小于预设阈值的编码区间进行重扩展；之后，根据重扩展后的编码区间，采用算术编码算法继续对数据中为完成的字符串进行编码，最后得到码值区间。

例如，当前数据包括 12 个字符串，预设阈值为 0.05，那么在采用算术编码算法对当前数据的 12 个字符串进行编码时，当编码区间小于 0.05 时，对数据的编码区间进行重扩展；例如，当第 7 个字符的编码区间为 0.1，第 8 个字符的编码区间为 0.04，那么，对 8 个字符的编码区间 0.04 进行重扩展，

例如，将0.04扩展为1或10等，之后，根据重扩展后的编码区间（1或10）采用算术编码算法继续对数据（第9至12字符）进行编码，得到码值区间（第12字符对应的编码区间）。

随着数据的字符串长度的增加，编码所得到的编码区间长度也会越来越小，在较小的编码区间长度中不容易取到合适的码值。因此，为了尽可能避免可能出现的无法正确进行算术编码的问题，本发明实施例通过对编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

应注意，进行区间扩展的所有信息将同样发送给解码器，即编码器将二进制码值传送给解码器的同时，也会将移位机制信息发送给解码器。用于同步信息，保证在解码的时候得到正确的解码结果。

图5是根据本发明另一实施例的数据处理的方法的示意性流程图。图5所示的方法可以由数据处理的设备执行，具体地可以由解码设备执行，如图5所示，该方法包括:

510，获取数据的码值和重扩展字符位置。

520，采用算术编码算法对数据的码值进行解码，得到解码区间。

530，根据重扩展字符位置，对数据的解码区间进行重缩放，得到重缩放的解码区间。

540，根据重缩放的解码区间，采用算术编码算法继续对数据进行解码，得到数据。

因此，本发明实施例通过对编码区间的重扩展的码值进行了解码码区间的重缩放，避免错误的解码，实现正确的解码。

具体地，作为另一实施例，在530中，根据重扩展字符位置，确定重缩放字符位置，其中重扩展字符位置与重缩放字符位置相倒置；根据重缩放字符位置对数据的解码区间进行重缩放，得到重缩放的解码区间：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。

应理解，重扩展字符位置与重缩放字符位置相倒置换句话说重扩展字符位置与重缩放字符位置相反（或相对称），例如，当前数据包括12个字符串，

对 12 个字符串中的第 10 个字符的编码区间重扩展，那么重扩展字符位置为第 10 个字符串的位置，根据第 10 个字符串的位置可以确定重缩放的字符位置为第 3 个字符的位置。

应理解，图 5 所示的数据处理方法与图 4 所示的数据处理方法相对应，区别在于，图 5 所示的解码过程为图 4 所示的编码过程的逆运算。图 5 的方法可由图 4 的逆过程而得到，为避免重复，此处不再赘述。

图 6 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 6 所示的方法可以由数据处理的设备执行，如图 6 所示，该方法包括:

610，采用算术编码算法对数据进行编码，得到码值区间。

620，根据码值区间获取码值。

630，存储码值。

640，根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种。

因此，本发明实施例通过对数据编码获取码值，并根据码值进行数据的等值比较、排列和模糊查询中的至少一个的应用，不同于现有的根据源数据进行数据的等值比较、排列和模糊查询，将原本复杂数据应用等效成采用码值的进行相应处理，快速简单。

可选地，作为另一实施例，数据为 ID 类字符串，应用操作包括等值比较，在 640 中，根据码值进行数据的等值比较;

进一步地，作为另一实施例，在 640 中，在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。

具体地，进行码值的等值比较可以理解为进行数据的匹配，例如，当两个码值相等时，可以确定该两个码值对应的数据为相同数据，也即匹配成功;当两个码值不相等时，可以确定该两个码值对应的数据为不同数据，也即匹配不成功。

可替代地，作为另一实施例，数据为 ID 类字符串或为字段的字母字符串，，应用操作包括排列，在 640 中，根据码值进行数据的排列。

进一步地，作为另一实施例，在 640 中，根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。

具体的，例如，进行数据的排序可以理解为对多个数据进行排序。例如，

有 5 个数据，对应 5 个码值，5 个码值由小到大排序，例如，当前码值为 5 个码值中的第 4 个,那么第 4 个码值对应的数据排在该 5 个数据中的第 4 位。

在现有的数据库实现中，通常会将压缩技术与查询操作分开考虑，即单独考虑数据存储技术以及相关的查询优化技术。数据库的重要作用在于存储记录所关心事物的重要描述，事物发展等相关信息。因此当事物描述较为冗繁时不仅占用大量的存储空间，也会带来查询上的不便。现有对数据在进行多种查询操作时(如字符串比较，字符串排序等)，因需要逐步比较字符串中的字符，会导致查询效率较低。而本发明实施例无需算术解码的过程，直接根据数据对应的码值进行数据的比较(匹配)和/或根据码值进行数据的排列，将原本复杂数据类型的查询操作等效成码值的查询，快速简单。

例如，当数据中只出现数字或字母的应用场景中(此时概率将全分给数字或全分给字母)。例如，数据为非拉丁字母文字。如中文、韩文、日文都有对应的音标。可以通过将对应文字转换成其音标表示方式，即只包含字母的字符串，在采用算术编码算法对数据进行编码，得到码值区间；在码值区间内存在与数据对应的码值时，确定码值的比特数小于数据的比特数；存储码值。根据码值对数据对应的字段进行排序。

例如，中文拼音字母出现概率分布(全拼)，A(0.107)，B(0.014)，C(0.017)，D(0.030)，E(0.062)，F(0.009)，G(0.060)，H(0.067)，I(0.141)，J(0.023)，K(0.008)，L(0.017)，M(0.014)，N(0.117)，O(0.065)，P(0.008)，Q(0.013)，R(0.006)，S(0.026)，T(0.015)，U(0.096)，V(0.001)，W(0.010)，X(0.020)，Y(0.028)，Z(0.026)。如图 2 所示，对于字段“优秀”“良好”“及格”，按字母顺序排列并编码，其中，“优秀”对应的字母字符串为“youxiu”，对应的码值为 0.96684845；“良好”对应的字母字符串为“lianghao”，对应的码值为 0.544375656；“及格”对应的字母字符串为“jige”，对应的码值为 0.516228。按照码值由小到大的依次排序为 0.516228、0.544375656 和 0.96684845，分别对应“及格”、“良好”和“优秀”。

可替代地，作为另一实施例，应用操作包括模糊查询，在 640 中，根据码值对数据进行模糊查询。

进一步地，作为另一实施例，在 640 中，根据码值是否在所需模糊查询的前缀字符串的编码区间中，确定数据是否包括前缀字符串，其中，在码值在所需模糊查询的前缀字符串的编码区间中时，数据包括前缀字符串，在码

值不在所需模糊查询的前缀字符串的编码区间中时，数据不包括前缀字符串。

换句话说，码值在所需模糊查询的前缀字符串的编码区间中时，数据满足模糊查询，码值不在所需模糊查询的前缀字符串的编码区间中时，数据不满足模糊查询。

具体而言，当字符串经过算术编码压缩后，得到的是一系列的码值。码值来自于对字符串编码后所得到的编码区间，而编码区间彼此之间相互不正交。我们还注意到字符串的编码区间总是包含于其前缀字符串的编码区间。例如，字符串'A12986572'的编码区间一定包含于'A1298'、'A12'等前缀字符串的编码区间中。只需判断码值是否在所需模糊查询的前缀字符串的编码区间中，就可进行模糊查询。例如，整数集空间{0,1,2,3}，出现概率分布{0.2, 0.5, 0.2, 0.1}。针对模糊查询%210xxx 而言，对于数据 212132、210312、210231 和 211123 进行该模糊查询:如图 3 所示，"210"的码值区间为[0.74，0.76]，212132、210312、210231 和 211123 对应的码值区间分别为 0.8238、0.7592、0.7576 和 0.7923，由于 0.7592 和 0.7576 在模糊查询的前缀字符串的编码区间[0.74，0.76]中，0.8238 和 0.7923 不在模糊查询的前缀字符串的编码区间[0.74，0.76]中，因此，210312 和 210231 满足模糊查询，212132 和 211123 不满足模糊查询。

因此，本发明实施例另外在模糊查询操作中，当索引字符数大于 2 时，即可抵消判断区间定位所需操作。

上文中结合图 1 至图 6 详细描述了本发明实施例的数据处理的方法，下面结合图 7 至图 11 的具体例子，更加详细地描述本发明实施例。应注意，图 7 至图 11 的例子仅仅是为了帮助本领域技术人员理解本发明实施例，而非要将本发明实施例限于所例示的具体数值或具体场景。本领域技术人员根据所给出的图 7 至图 11 的例子，显然可以进行各种等价的修改或变化，这样的修改或变化也落入本发明实施例的范围内。

图 7 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 7 中的方法中显示了根据收益来评估是否采用算术编码的过程，如图 7 所示的方法包括:

710，概率模型估计。

720，输入序列。

具体地，该序列可以为源数据的字符串序列，例如，可以为数字与字母特殊组合的数据类型。或者仅数字或仅字母的数据类型，例如，输入的序列可以为数据单号、图书索引号、公交车牌号、产品单据号、IMSI 或 IMEI 等，以及中文、韩文、日文对应的音标等。

730 编码区间。

具体地，根据算术编码算法，确定源数据的编码区间。

740，码值选取。

具体地，确定是否在编码区间中存在与数据对应的码值，换句话说，确定是否可以选取码值，如果是则进行步骤 750，如果无法选取码值则进行 790。

750，收益评估。

具体地，在码值区间内存在与数据对应的码值时，确定码值的比特数是否满足要求，例如，确定码值的比特数是否小于数据的比特数。

760，是否满足要求。

在满足要求时进行步骤 770，否则，在不满足要求时进行步骤 780。

770，应用。

具体地，可以根据码值进行等值比较、排序以及模糊查询等操作。

780，放弃。

具体地，放弃算术编码。

790，放弃。

具体地，放弃算术编码。

具体地，本发明实施例可判断在最后所得到的编码区间内是否可以取到合适的码值，如果无法取到将直接存储数据而无需算术编码。当可以取到码值时，将判断为了表示码值是否会导致所需的比特位数多于表示之前原数据所需要的比特位数，如果无压缩收益将放弃采用算术编码。

因此，本发明实施例通过采用算术编码算法对数据进行编码，得到码值区间。在码值区间内存在与数据对应的码值时，确定码值的比特数小于数据的比特数，并存储码值，本发明实施例进行收益判断，能够降低错误的编码，减小数据的存储空间。

图 8 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 8 中的方法中显示了基于编码区间重扩展的算术编码和算术解码的过程，如图 8 所示的方法包括:

810，输入源序列。

具体地，该序列可以为源数据的字符串序列，例如，可以为数字与字母特殊组合的数据类型。或者仅数字或仅字母的数据类型，例如，输入的序列可以为数据单号、图书索引号、公交车牌号、产品单据号、IMSI 或 IMEI 等，以及中文、韩文、日文对应的音标等。

820，源模型。

具体地，源模型中包括各个字符的概率值。数据处理的设备可以根据源模型对源序列进行算术编码。

830，算术编码。

具体地，采用算术编码算法根据源模型对数据的字符串逐步进行编码。

840，是否满足预设条件。

具体地，判断当已完成编码的数据的字符个数是否达到预设字符个数阈值或者数据的编码区间的长度是否小于预设的阈值。如果是则执行步骤 850，否则执行步骤 840。

850，编码区间重扩展

在满足预设条件时，对数据的编码区间进行重扩展；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。最后得到源序列的码值。

例如，当前数据包括 12 个字符串，预设阈值为 0.05，那么在采用算术编码算法对当前数据的 12 个字符串进行编码时，当编码区间小于 0.05 时，对数据的编码区间进行重扩展；例如，当第 7 个字符的编码区间为 0.1，第 8 个字符的编码区间为 0.04，那么，对 8 个字符的编码区间 0.04 进行重扩展，例如，将 0.04 扩展为 1 或 10 等，之后，根据重扩展后的编码区间（1 或 10）采用算术编码算法继续对数据（第 9 至 12 字符）进行编码，得到码值区间（第 12 字符对应的编码区间）。

再例如，当前数据包括 12 个字符串，预设阈值为 0.05，那么在采用算术编码算法对当前数据的 12 个字符串进行编码时，当编码区间小于 0.05 时，对数据的编码区间进行重扩展；例如，当第 7 个字符的编码区间为 0.1，第 8 个字符的编码区间为 0.04，那么，对 8 个字符的编码区间 0.04 进行重扩展，例如，将 0.04 扩展为 1 或 10 等，之后，根据重扩展后的编码区间（1 或 10）采用算术编码算法继续对数据（第 9 至 12 字符）进行编码，得到码值区间

（第 12 字符对应的编码区间）。

860，算术解码。

具体地，根据源模型，以及编码的信息，例如，如果在解码时进行了区间扩展，同样在解码时进行相反的操作，进行区间的缩放。

应注意，进行区间扩展的所有信息将同样发送给解码器，即编码器将二进制码值传送给解码器的同时，也会将移位机制信息发送给解码器。用于同步信息，保证在解码的时候得到正确的解码结果。

870，源模型。

与 820 中的源模型对应，具体地，源模型中包括各个字符的概率值。解码数据的设备可以根据源模型对源序列进行算术解码。

880，得到解码序列。

具体地，解码序列可以与源序列相同。

随着数据的字符串长度的增加，编码所得到的编码区间长度也会越来越小，在较小的编码区间长度中不容易取到合适的码值。因此，为了尽可能避免可能出现的无法正确进行算术编码的问题，能够容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

图 9 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 9 中的方法中显示了 ID 类字符串的算术编码及查询操作的过程，如图 9 所示的方法包括:

910，获取 ID 类字符串。

具体地，ID 类字符串可以为数据单号、图书索引号、公交车牌号等字字母出现的概率较低，而数字出现的概率较高的字符串。比如通常来说字母只占据字符串中的 1 位或者 2 位。

920，设计概论模型。

具体地，概论模型包括各个字符的概率。

930，算术编码。

具体地，采用算术编码算法根据概论模型对数据的字符串逐步进行编码。

940，评估判断。

具体地，确定是否在编码区间中存在与数据对应的码值，换句话说，确

定是否可以选取码值，如果是，则确定码值的比特数是否满足要求，例如，确定码值的比特数是否小于数据的比特数。

950，是否满足要求。

如果满足要求，则进行 970，否则，进行步骤 960。

960，放弃。

970，获取码值。

具体地，从码值区间中确定与 ID 类字符串对应的码值。

980，等值比较和/或排序。

具体地，可以根据码值进行数据的等值比较操作，在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。具体地，进行码值的等值比较可以理解为进行数据的匹配，例如，当两个码值相等时，可以确定该两个码值对应的数据为相同数据，也即匹配成功；当两个码值不相等时，可以确定该两个码值对应的数据为不同数据，也即匹配不成功。

还可以据码值进行数据的排序，例如，根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。具体地，进行数据的排序可以理解为对多个数据进行排序。例如，有 5 个数据，对应 5 个码值，5 个码值由小到大排序，例如，当前码值为 5 个码值中的第 4 个，那么第 4 个码值对应的数据排在该 5 个数据中的第 4 位。

图 10 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 10 中的方法中显示了字段的字母字符串的算术编码及字段排序的过程，如图 10 所示的方法包括：

1010，获取字段。

具体地，该字段可以为中文、韩文、日文字段等，本发明实施例并不限于此，该字段还可以为通过音标转换成字母字符串的其他字段。例如，该字段可以为中文字段“优秀”“良好”和“及格”。

1020，字母字符串。

具体地，将字段转换成字母字符串。例如，“优秀”、“良好”和“及格”对应的字母字符串分别为“youxiu”、“lianghao”和“jige”。

1030，拼音字母概率。

具体地，获取各个字母的概率，例如，中文拼音字母出现概率分布(全拼)，A(0.107)，B(0.014)，C(0.017)，D(0.030)，E(0.062)，F(0.009)，G(0.060)，

H(0.067), I(0.141), J(0.023), K(0.008), L(0.017), M(0.014), N(0.117), O(0.065), P(0.008), Q(0.013), R(0.006), S(0.026), T(0.015), U(0.096), V(0.001), W(0.010), X(0.020), Y(0.028), Z(0.026)。

1040，算术编码。

具体地，根据上述拼音字母概率，采用算术编码算法进行编码。

1050，评估判断。

具体地，确定是否在编码区间中存在与数据对应的码值，换句话说，确定是否可以选取码值，如果是，则确定码值的比特数是否满足要求，例如，确定码值的比特数是否小于数据的比特数。

1060，是否满足要求。

如果满足要求，则进行 1080，否则，进行步骤 1070。

1070，返回。

1080，获取码值。

具体地，从码值区间中确定与字母字符串对应的码值。

1090，字段排序。

例如，对于字段“优秀”“良好”“及格”，按字母顺序排列并编码，其中，“优秀”对应的字母字符串为“youxiu”，对应的码值为 0.96684845；“良好”对应的字母字符串为“lianghao”，对应的码值为 0.544375656；“及格”对应的字母字符串为“jige”，对应的码值为 0.516228。按照码值由小到大的依次排序为 0.516228、0.544375656 和 0.96684845，分别对应“及格”、“良好”和“优秀”。

图 11 是根据本发明另一实施例的数据处理的方法的示意性流程图。图 11 中的方法中显示了基于算术编码的模糊查询的过程，如图 11 所示的方法包括:

1110，选取索引片段。

具体地，获取需要模糊查询的前缀字符串片段。例如，该索引片段为“210”。

1120，算术编码。

具体地，根据算术编码算法对索引片段进行编码。

1130，码值区间。

具体地，获取索引片段的码值区间。例如，整数集空间{0,1,2,3}，出现概率分布{0.2, 0.5, 0.2, 0.1}。针对“210”而言，“210”的码值区间为[0.74, 0.76]。

1140，获取序列对应的码值。

具体地，获取需要模糊查询的序列对应的码值，例如，212132、210312、210231 和 211123 对应的码值区间分别为 0.8238、0.7592、0.7576 和 0.7923。

1150，检验并记录。

具体地，根据序列对应的码值和索引片段的码值区间进行模糊查询并记录结果。例如，210”的码值区间为[0.74，0.76]，212132、210312、210231 和 211123 对应的码值区间分别为 0.8238、0.7592、0.7576 和 0.7923，由于 0.7592 和 0.7576 在模糊查询的前缀字符串的编码区间[0.74，0.76]中，0.8238 和 0.7923 不在模糊查询的前缀字符串的编码区间[0.74，0.76]中，因此，210312 和 210231 满足模糊查询条件，212132 和 211123 不满足模糊查询条件。

1160，是否结束。

具体地，如果是，则进行步骤 870，否则进行步骤 840 获取另一个序列对应的码字。

1170，输出结果。

应注意，图 7 至图 11 的例子是为了帮助本领域技术人员更好地理解本发明实施例，而非要限制本发明实施例的范围。本领域技术人员根据所给出的图 7 至图 11 的例子，显然可以进行各种等价的修改或变化，这样的修改或变化也落入本发明实施例的范围内。

应理解，上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

上文中，结合图 1 至图 11 详细描述了本发明实施例的数据处理的方法，下面将结合图 12 至图 19 描述本发明实施例的数据处理的设备。

图 12 是根据本发明一个实施例的数据处理的设备的示意框图。如图 12 所示的数据处理的设备 1200 可以是编码设备，如图 12 所示的数据处理的设备 1200 包括：编码单元 1210、获取单元 1220、比较单元 1230 和第一存储单元 1240。

具体地，编码单元 1210，用于采用算术编码算法对数据进行编码，得到码值区间；获取单元 1220，用于在码值区间内存在与数据对应的码值时，根据码值区间获取码值；比较单元 1230，用于将码值的比特数与数据的比特数的进行比较，以获取比较结果；第一存储单元 1240，用于根据比较结果进行

存储操作。

因此，本发明实施例通过采用算术编码算法对数据进行编码，得到码值区间。在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；根据比较结果进行存储操作。本发明实施例能够降低错误的编码，减小数据的存储空间。

可选地，作为另一实施例，比较结果为码值的比特数小于数据的比特数，其中，第一存储单元根据比较结果，存储码值。

可替代地，作为另一实施例，比较结果为码值的比特数大于或等于数据的比特数，其中，第一存储单元根据比较结果，存储数据。

可选地，作为另一实施例，该设备还包括：应用单元，用于根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种。

可选地，作为另一实施例，数据为 ID 类字符串，应用操作包括等值比较，应用单元根据码值进行数据的等值比较。

具体地，作为另一实施例，应用单元在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。

可替代地，作为另一实施例，数据为 ID 类字符串或为字段的字母字符串，应用操作包括排列，应用单元根据码值进行数据的排列。

具体地，作为另一实施例，应用单元根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。

可替代地，作为另一实施例，应用操作包括模糊查询，应用单元根据码值对数据进行模糊查询。

具体地，作为另一实施例，应用单元根据码值是否在所需模糊查询的前缀字符串的编码区间中，确定数据是否包括前缀字符串，其中，在码值在所需模糊查询的前缀字符串的编码区间中时，数据包括前缀字符串，在码值不在所需模糊查询的前缀字符串的编码区间中时，数据不包括前缀字符串。

可选地，作为另一实施例，编码单元 1210 采用算术编码算法对数据进行编码，得到编码区间；对数据的编码区间进行重扩展，得到重扩展后的编码区间；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

进一步地，编码单元 1210 当满足以下条件中的至少一个时，对数据的

编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。

可选地，作为另一实施例，该设备还包括：确定单元，用于在获取单元1220获取码值之前，确定码值区间内是否存在与数据对应的合适的码值。

可选地，作为另一实施例，该设备还包括：第二存储单元，用于在码值区间内不存在与数据对应的合适的码值时，存储数据。

应理解，图12所示的数据处理的设备与图1所示的数据处理的方法相对应，图12所示的数据处理的设备能够实现图1的数据处理的方法的各个过程，图12所示的数据处理的设备的功能可参见图1数据处理的方法的相关描述，为避免重复，此处不再赘述。

图13是根据本发明另一实施例的数据处理的设备的示意框图。如图13所示的数据处理的设备1300可以为编码设备，如图13所示的设备1300包括：第一编码单元1310、扩展单元1320、第二编码单元1330、获取单元1340和存储单元1350。

具体地，第一编码单元1310用于采用算术编码算法对数据进行编码，得到编码区间；扩展单元1320用于对数据的编码区间进行重扩展，得到重扩展后的编码区间；第二编码单元1330用于根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。获取单元1340用于根据码值区间获取码值；存储单元1350用于存储码值。

因此，本发明实施例通过对数据的编码区间进行编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

可选地，作为另一实施例，扩展单元1320当满足以下条件中的至少一个时，对数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。

应理解，图13所示的数据处理的设备与图4所示的数据处理的方法相对应，图13所示的数据处理的设备能够实现图4的数据处理的方法的各个

过程，图 13 所示的数据处理的设备的功能可参见图 4 数据处理的方法的相关描述，为避免重复，此处不再赘述。

图 14 是根据本发明另一实施例的数据处理的设备的示意框图。如图 14 所示的数据处理的设备 1400 可以为解码设备，如图 14 所示的设备 1400 包括：第一获取单元 1410、第一解码单元 1420、缩放单元 1430 和第二解码单元 1440。

具体地，第一获取单元 1410 用于获取数据的码值和重扩展字符位置；第一解码单元 1420 用于采用算术编码算法对数据的码值进行解码，得到解码区间；缩放单元 1430 用于根据重扩展字符位置，对数据的解码区间进行重缩放，得到重缩放的解码区间；第二解码单元 1440 用于根据重缩放的解码区间，采用算术编码算法继续对数据进行解码，得到数据。

因此，本发明实施例通过对编码区间的重扩展的码值进行了解码码区间的重缩放，避免错误的解码，实现正确的解码。

可选地，作为另一实施例，缩放单元 1430 根据重扩展字符位置，确定重缩放字符位置，其中重扩展字符位置与重缩放字符位置相倒置；根据重缩放字符位置对数据的解码区间进行重缩放，得到重缩放的解码区间。

应理解，图 14 所示的数据处理的设备与图 5 所示的数据处理的方法相对应，图 14 所示的数据处理的设备能够实现图 5 的数据处理的方法的各个过程，图 14 所示的数据处理的设备的功能可参见图 5 数据处理的方法的相关描述，为避免重复，此处不再赘述。

图 15 是根据本发明另一实施例的数据处理的设备的示意框图。如图 15 所示的数据处理的设备 1500 可以为编码设备，如图 15 所示的设备 1500 包括：编码单元 1510、获取单元 1520、存储单元 1530 和应用单元 1540。

具体地，编码单元 1510 用于采用算术编码算法对数据进行编码，得到码值区间；获取单元 1520 用于根据码值区间获取码值；存储单元 1530 用于存储码值；应用单元 1540 用于根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种。

因此，本发明实施例通过对数据编码获取码值，并根据码值进行数据的等值比较、排列和模糊查询中的至少一个的应用，不同于现有的根据源数据进行数据的等值比较、排列和模糊查询，将原本复杂数据应用等效成采用码值的进行相应处理，快速简单。

可选地，作为另一实施例，数据为 ID 类字符串，应用操作包括等值比较，应用单元 1540 根据码值进行数据的等值比较；

具体地，作为另一实施例，应用单元 1540 在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。

可替代地，作为另一实施例，数据为 ID 类字符串或为字段的字母字符串，应用操作包括排列，应用单元 1540 根据码值进行数据的排列。

具体地，作为另一实施例，应用单元 1540 根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。

可替代地，作为另一实施例，应用操作包括模糊查询，应用单元 1540 根据码值对数据进行模糊查询。

具体地，作为另一实施例，应用单元 1540 根据码值是否在所需模糊查询的前缀字符串的编码区间中，确定数据是否包括前缀字符串，其中，在码值在所需模糊查询的前缀字符串的编码区间中时，数据包括前缀字符串，在码值不在所需模糊查询的前缀字符串的编码区间中时，数据不包括前缀字符串。

应理解，图 15 所示的数据处理的设备与图 6 所示的数据处理的方法相对应，图 15 所示的数据处理的设备能够实现图 6 的数据处理的方法的各个过程，图 15 所示的数据处理的设备的功能可参见图 6 数据处理的方法的相关描述，为避免重复，此处不再赘述。

图 16 是根据本发明另一实施例的数据处理的设备的示意框图。如图 16 所示的数据处理的设备 1600 可以是编码设备，如图 16 所示的数据处理的设备 1600 包括：包括处理器 1610、存储器 1620 和总线系统 1630。

具体地，处理器 1610 通过总线系统 1630 调用存储在存储器 1620 中的代码，采用算术编码算法对数据进行编码，得到码值区间；在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；根据比较结果进行存储操作。

本发明实施例通过采用算术编码算法对数据进行编码，得到码值区间。在码值区间内存在与数据对应的码值时，根据码值区间获取码值；将码值的比特数与数据的比特数的进行比较，以获取比较结果；根据比较结果进行存储操作。本发明实施例能够降低错误的编码，减小数据的存储空间。

上述本发明实施例揭示的方法可以应用于处理器 1610 中，或者由处理器 1610 实现。处理器 1610 可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器 1610 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 1610 可以是通用处理器、数字信号处理器（英文 Digital Signal Processor，简称 DSP）、专用集成电路（英文 Application Specific Integrated Circuit，简称 ASIC）、现成可编程门阵列（英文 Field Programmable Gate Array，简称 FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器（英文 Random Access Memory，简称 RAM）、闪存、只读存储器（英文 Read-Only Memory，简称 ROM）、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 1620，处理器 1610 读取存储器 1620 中的信息，结合其硬件完成上述方法的步骤，该总线系统 1630 除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统 1630。

可选地，作为另一实施例，比较结果为码值的比特数小于数据的比特数，其中，处理器 1610 根据比较结果，存储码值。

可替代地，作为另一实施例，比较结果为码值的比特数大于或等于数据的比特数，其中，处理器 1610 根据比较结果，存储数据。

可选地，作为另一实施例，该设备还包括：处理器 1610 用于根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种。

可选地，作为另一实施例，数据为 ID 类字符串，应用操作包括等值比较，处理器 1610 根据码值进行数据的等值比较。

具体地，作为另一实施例，处理器 1610 在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。

可替代地，作为另一实施例，数据为 ID 类字符串或为字段的字母字符串，应用操作包括排列，处理器 1610 根据码值进行数据的排列。

具体地，作为另一实施例，处理器 1610 根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。

可替代地，作为另一实施例，应用操作包括模糊查询，处理器 1610 根据码值对数据进行模糊查询。

具体地，作为另一实施例，处理器 1610 根据码值是否在所需模糊查询的前缀字符串的编码区间中确定数据是否包括前缀字符串，其中，在码值在所需模糊查询的前缀字符串的编码区间中时，数据包括前缀字符串，在码值不在所需模糊查询的前缀字符串的编码区间中时，数据不包括前缀字符串。

可选地，作为另一实施例，处理器 1610 采用算术编码算法对数据进行编码，得到编码区间；对数据的编码区间进行重扩展，得到重扩展后的编码区间；根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间。

进一步地，处理器 1610 当满足以下条件中的至少一个时，对数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。

可选地，作为另一实施例，处理器 1610 获取码值之前，确定码值区间内是否存在与数据对应的合适的码值。

可选地，作为另一实施例，处理器 1610 在码值区间内不存在与数据对应的合适的码值时，存储数据。

应理解，图 16 所示的数据处理的设备与图 1 所示的数据处理的方法相对应，图 16 所示的数据处理的设备能够实现图 1 的数据处理的方法的各个过程，图 16 所示的数据处理的设备的功能可参见图 1 数据处理的方法的相关描述，为避免重复，此处不再赘述。

图 17 是根据本发明另一实施例的数据处理的设备的示意框图。如图 17 所示的数据处理的设备 1700 可以为编码设备，如图 17 所示的设备 1700 包括：包括处理器 1710、存储器 1720 和总线系统 1730。

具体地，处理器 1710 通过总线系统 1730 调用存储在存储器 1720 中的代码，采用算术编码算法对数据进行编码，得到编码区间；对数据的编码区间进行重扩展，得到重扩展后的编码区间；根据重扩展后的编码区间，采用

算术编码算法继续对数据进行编码，得到码值区间。根据码值区间获取码值；存储码值。

因此，本发明实施例通过对数据的编码区间进行编码区间的重扩展，由于编码区间进行了重扩展进而使得码值区间同样的扩大，因此本发明实施例能够在扩大的码值区间中更容易取到合适的码值，避免错误的编码，实现正确的编码。另外，本发明实施例进行区间的重扩展能够实现使用有限位数的空间来表示足够长的字符串数据。

上述本发明实施例揭示的方法可以应用于处理器 1710 中，或者由处理器 1710 实现。处理器 1710 可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器 1710 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 1710 可以是通用处理器、数字信号处理器（英文 Digital Signal Processor，简称 DSP）、专用集成电路（英文 Application Specific Integrated Circuit，简称 ASIC）、现成可编程门阵列（英文 Field Programmable Gate Array，简称 FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器（英文 Random Access Memory，简称 RAM）、闪存、只读存储器（英文 Read-Only Memory，简称 ROM）、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 1720，处理器 1710 读取存储器 1720 中的信息，结合其硬件完成上述方法的步骤，该总线系统 1730 除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统 1730。

可选地，作为另一实施例，处理器 1710 当满足以下条件中的至少一个时，对数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置：已完成编码的数据的字符个数达到字符个数阈值，和数据的编码区间的长度小于区间阈值。

应理解，图 17 所示的数据处理的设备与图 4 所示的数据处理的方法相对应，图 17 所示的数据处理的设备能够实现图 4 的数据处理的方法的各个

过程，图 17 所示的数据处理的设备的功能可参见图 4 数据处理的方法的相关描述，为避免重复，此处不再赘述。

图 18 是根据本发明另一实施例的数据处理的设备的示意框图。如图 18 所示的数据处理的设备 1800 可以为解码设备，如图 18 所示的设备 1800 包括：包括处理器 1810、存储器 1820 和总线系统 1830。

具体地，处理器 1810 通过总线系统 1830 调用存储在存储器 1820 中的代码，获取数据的码值和重扩展字符位置；采用算术编码算法对数据的码值进行解码，得到解码区间；根据重扩展字符位置，对数据的解码区间进行重缩放，得到重缩放的解码区间；根据重缩放的解码区间，采用算术编码算法继续对数据进行解码，得到数据。

因此，本发明实施例通过对编码区间的重扩展的码值进行了解码码区间的重缩放，避免错误的解码，实现正确的解码。

上述本发明实施例揭示的方法可以应用于处理器 1810 中，或者由处理器 1810 实现。处理器 1810 可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器 1810 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 1810 可以是通用处理器、数字信号处理器（英文 Digital Signal Processor，简称 DSP）、专用集成电路（英文 Application Specific Integrated Circuit，简称 ASIC）、现成可编程门阵列（英文 Field Programmable Gate Array，简称 FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器（英文 Random Access Memory，简称 RAM）、闪存、只读存储器（英文 Read-Only Memory，简称 ROM）、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 1820，处理器 1810 读取存储器 1820 中的信息，结合其硬件完成上述方法的步骤，该总线系统 1830 除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统 1830。

可选地，作为另一实施例，处理器 1810 根据重扩展字符位置，确定重

缩放字符位置，其中重扩展字符位置与重缩放字符位置相倒置；根据重缩放字符位置对数据的解码区间进行重缩放，得到重缩放的解码区间。

应理解，图 18 所示的数据处理的设备与图 5 所示的数据处理的方法相对应，图 18 所示的数据处理的设备能够实现图 5 的数据处理的方法的各个过程，图 18 所示的数据处理的设备的功能可参见图 5 数据处理的方法的相关描述，为避免重复，此处不再赘述。

图 19 是根据本发明另一实施例的数据处理的设备的示意框图。如图 19 所示的数据处理的设备 1900 可以为编码设备，如图 19 所示的设备 1900 包括：包括处理器 1910、存储器 1920 和总线系统 1930。

具体地，处理器 1910 通过总线系统 1930 调用存储在存储器 1920 中的代码，采用算术编码算法对数据进行编码，得到码值区间；根据码值区间获取码值；存储码值；根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种。

因此，本发明实施例通过对数据编码获取码值，并根据码值进行数据的等值比较、排列和模糊查询中的至少一个的应用，不同于现有的根据源数据进行数据的等值比较、排列和模糊查询，将原本复杂数据应用等效成采用码值的进行相应处理，快速简单。

上述本发明实施例揭示的方法可以应用于处理器 1910 中，或者由处理器 1910 实现。处理器 1910 可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器 1910 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 1910 可以是通用处理器、数字信号处理器（英文 Digital Signal Processor，简称 DSP）、专用集成电路（英文 Application Specific Integrated Circuit，简称 ASIC）、现成可编程门阵列（英文 Field Programmable Gate Array，简称 FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器（英文 Random Access Memory，简称 RAM）、闪存、只读存储器（英文 Read-Only Memory，简称 ROM）、可编程只读存储器或者电可擦写可编程存储器、寄

存器等本领域成熟的存储介质中。该存储介质位于存储器 1920，处理器 1910 读取存储器 1920 中的信息，结合其硬件完成上述方法的步骤，该总线系统 1930 除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统 1930。

可选地，作为另一实施例，数据为 ID 类字符串，应用操作包括等值比较，处理器 1910 根据码值进行数据的等值比较；

具体地，作为另一实施例，处理器 1910 在码值与待比较的码值相等时，确定数据与待比较的码值对应的数据为相同数据。

可替代地，作为另一实施例，数据为 ID 类字符串或为字段的字母字符串，应用操作包括排列，处理器 1910 根据码值进行数据的排列。

具体地，作为另一实施例，处理器 1910 根据码值的大小，确定码值在待排列码值中的位置，码值的位置用于表示数据在与待排列的码值对应的数据中的位置。

可替代地，作为另一实施例，应用操作包括模糊查询，处理器 1910 根据码值对数据进行模糊查询。

具体地，作为另一实施例，处理器 1910 根据码值是否在所需模糊查询的前缀字符串的编码区间中，确定数据是否包括前缀字符串，其中，在码值在所需模糊查询的前缀字符串的编码区间中时，数据包括前缀字符串，在码值不在所需模糊查询的前缀字符串的编码区间中时，数据不包括前缀字符串。

应理解，图 19 所示的数据处理的设备与图 6 所示的数据处理的方法相对应，图 19 所示的数据处理的设备能够实现图 6 的数据处理的方法的各个过程，图 19 所示的数据处理的设备的功能可参见图 6 数据处理的方法的相关描述，为避免重复，此处不再赘述。

应理解，说明书通篇中提到的“一个实施例”或“一实施例”意味着与实施例有关的特定特征、结构或特性包括在本发明的至少一个实施例中。因此，在整个说明书各处出现的“在一个实施例中”或“在一实施例中”未必一定指相同的实施例。此外，这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。应理解，在本发明的各种实施例中，上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

另外，本文中术语“系统”和“网络”在本文中常被可互换使用。本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

应理解，在本发明实施例中，“与 A 相应的 B”表示 B 与 A 相关联，根据 A 可以确定 B。但还应理解，根据 A 确定 B 并不意味着仅仅根据 A 确定 B，还可以根据 A 和/或其它信息确定 B。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、装置或单元的间接耦合或通信连接，也可以是电的，机械的或其它的形式连接。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本发明实施例方案的目的。

另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以是两个或两个以上单元集成在

一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本发明可以用硬件实现，或固件实现，或它们的组合方式来实现。当使用软件实现时，可以将上述功能存储在计算机可读介质中或作为计算机可读介质上的一个或多个指令或代码进行传输。计算机可读介质包括计算机存储介质和通信介质，其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是计算机能够存取的任何可用介质。以此为例但不限于：计算机可读介质可以包括 RAM、ROM、EEPROM、CD-ROM 或其他光盘存储、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质。此外。任何连接可以适当的成为计算机可读介质。例如，如果软件是使用同轴电缆、光纤光缆、双绞线、数字用户线（DSL）或者诸如红外线、无线电和微波之类的无线技术从网站、服务器或者其他远程源传输的，那么同轴电缆、光纤光缆、双绞线、DSL 或者诸如红外线、无线和微波之类的无线技术包括在所属介质的定影中。如本发明所使用的，盘（Disk）和碟（disc）包括压缩光碟（CD）、激光碟、光碟、数字通用光碟（DVD）、软盘和蓝光光碟，其中盘通常磁性的复制数据，而碟则用激光来光学的复制数据。上面的组合也应当包括在计算机可读介质的保护范围之内。

总之，以上所述仅为本发明技术方案的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求

1. 一种数据处理的方法，其特征在于，包括:

采用算术编码算法对数据进行编码，得到码值区间;

在所述码值区间内存在与所述数据对应的码值时，根据所述码值区间获取所述码值;

将所述码值的比特数与所述数据的比特数的进行比较，以获取比较结果;

根据所述比较结果进行存储操作。

2. 根据权利要求 1 所述的方法，其特征在于，所述比较结果为所述码值的比特数小于所述数据的比特数，

其中，所述根据所述比较结果进行存储操作，包括:

根据所述比较结果，存储所述码值。

3. 根据权利要求 1 所述的方法，其特征在于，所述比较结果为所述码值的比特数大于或等于所述数据的比特数，

其中，所述根据所述比较结果进行存储操作，包括:

根据所述比较结果，存储所述数据。

4. 根据权利要求 2 所述的方法，其特征在于，还包括:

根据所述码值进行所述数据的应用操作，所述应用操作包括等值比较、排列和模糊查询中的至少一种。

5. 根据权利要求 4 所述的方法，其特征在于，所述数据为标识 ID 类字符串，所述应用操作包括等值比较，所述根据所述码值进行所述数据的应用操作，包括:

在所述码值与待比较的码值相等时，确定所述数据与所述待比较的码值对应的数据为相同数据。

6. 根据权利要求 4 所述的方法，其特征在于，所述数据为 ID 类字符串或为字段的字母字符串，所述应用操作包括排列，所述根据所述码值进行所述数据的应用操作，包括:

根据所述码值的大小，确定所述码值在待排列码值中的位置，所述码值的位置用于表示所述数据在与所述待排列的码值对应的数据中的位置。

7. 根据权利要求 4 所述的方法，其特征在于，所述应用操作包括模糊查询，所述根据所述码值进行所述数据的应用操作，包括:

根据所述码值是否在所需模糊查询的前缀字符串的编码区间中，确定所述数据是否包括所述前缀字符串，

其中，在所述码值在所需模糊查询的前缀字符串的编码区间中时，所述数据包括所述前缀字符串，

在所述码值不在所需模糊查询的前缀字符串的编码区间中时，所述数据不包括所述前缀字符串。

8. 根据权利要求 1 至 7 中任一项所述的方法，其特征在于，所述采用算术编码算法对数据进行编码，得到码值区间，包括:

采用算术编码算法对所述数据进行编码，得到编码区间;

对所述数据的编码区间进行重扩展，得到重扩展后的编码区间;

根据所述重扩展后的编码区间，采用算术编码算法继续对所述数据进行编码，得到码值区间。

9. 根据权利要求 8 所述的方法，其特征在于，所述对所述数据的编码区间进行重扩展，得到重扩展后的编码区间，包括:

当满足以下条件中的至少一个时，对所述数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置:

已完成编码的所述数据的字符个数达到字符个数阈值，和所述数据的编码区间的长度小于区间阈值。

10. 根据权利要求 1 至 9 中任一项所述的方法，其特征在于，在所述码值区间内存在与所述数据对应的码值时，根据所述码值区间获取所述码值之前，还包括:

确定所述码值区间内是否存在与所述数据对应的合适的码值。

11. 根据权利要求 10 所述的方法，其特征在于，还包括:

在所述码值区间内不存在与所述数据对应的合适的码值时，存储所述数据。

12. 一种数据处理的方法，其特征在于，包括:

采用算术编码算法对数据进行编码，得到码值区间;

根据所述码值区间获取所述码值;

存储所述码值;

根据所述码值进行所述数据的应用操作，所述应用操作包括等值比较、排列和模糊查询中的至少一种。

13. 根据权利要求 12 所述的方法，其特征在于，所述数据为 ID 类字符串，所述应用操作包括等值比较，所述根据所述码值进行所述数据的应用操作，包括:

在所述码值与待比较的码值相等时，确定所述数据与所述待比较的码值对应的数据为相同数据。

14. 根据权利要求 12 所述的方法，其特征在于，所述数据为 ID 类字符串或为字段的字母字符串，所述应用操作包括排列，所述根据所述码值进行所述数据的应用操作，包括:

根据所述码值的大小，确定所述码值在待排列码值中的位置，所述码值的位置用于表示所述数据在与所述待排列的码值对应的数据中的位置。

15. 根据权利要求 12 所述的方法，其特征在于，所述应用操作包括模糊查询，所述根据所述码值进行所述数据的应用操作，包括:

根据所述码值是否在所需模糊查询的前缀字符串的编码区间中，确定所述数据是否包括所述前缀字符串，

其中，在所述码值在所需模糊查询的前缀字符串的编码区间中时，所述数据包括所述前缀字符串，

在所述码值不在所需模糊查询的前缀字符串的编码区间中时，所述数据不包括所述前缀字符串。

16. 一种数据处理的设备，其特征在于，包括:

编码单元，用于采用算术编码算法对数据进行编码，得到码值区间;

获取单元，用于在所述码值区间内存在与所述数据对应的码值时，根据所述码值区间获取所述码值;

比较单元，用于将所述码值的比特数与所述数据的比特数的进行比较，以获取比较结果;

第一存储单元，用于根据所述比较结果进行存储操作。

17. 根据权利要求 16 所述的设备，其特征在于，所述比较结果为所述码值的比特数小于所述数据的比特数，

其中，所述第一存储单元根据所述比较结果，存储所述码值。

18. 根据权利要求 16 所述的设备，其特征在于，所述比较结果为所述码值的比特数大于或等于所述数据的比特数，

其中，所述第一存储单元根据所述比较结果，存储所述数据。

19. 根据权利要求 17 所述的设备，其特征在于，还包括:

应用单元，用于根据所述码值进行所述数据的应用操作，所述应用操作包括等值比较、排列和模糊查询中的至少一种。

20. 根据权利要求 19 所述的设备，其特征在于，所述数据为 ID 类字符串，所述应用操作包括等值比较，所述应用单元在所述码值与待比较的码值相等时，确定所述数据与所述待比较的码值对应的数据为相同数据。

21. 根据权利要求 19 所述的设备，其特征在于，所述数据为 ID 类字符串或为字段的字母字符串，所述应用操作包括排列，所述应用单元根据所述码值的大小，确定所述码值在待排列码值中的位置，所述码值的位置用于表示所述数据在与所述待排列的码值对应的数据中的位置。

22. 根据权利要求 19 所述的设备，其特征在于，所述应用操作包括模糊查询，所述应用单元根据所述码值是否在所需模糊查询的前缀字符串的编码区间中确定所述数据是否包括所述前缀字符串，

其中，在所述码值在所需模糊查询的前缀字符串的编码区间中时，所述数据包括所述前缀字符串，

在所述码值不在所需模糊查询的前缀字符串的编码区间中时，所述数据不包括所述前缀字符串。

23. 根据权利要求 16 至 22 中任一项所述的设备，其特征在于，所述编码单元采用算术编码算法对所述数据进行编码，得到编码区间；对所述数据的编码区间进行重扩展，得到重扩展后的编码区间；根据所述重扩展后的编码区间，采用算术编码算法继续对所述数据进行编码，得到码值区间。

24. 根据权利要求 23 所述的设备，其特征在于，所述编码单元当满足以下条件中的至少一个时，对所述数据的编码区间进行重扩展，得到重扩展后的编码区间，并记录相应重扩展字符位置:

已完成编码的所述数据的字符个数达到字符个数阈值，和所述数据的编码区间的长度小于区间阈值。

25. 根据权利要求 26 至 24 中任一项所述的设备，其特征在于，还包括:

确定单元，用于在所述获取单元获取所述码值之前，确定所述码值区间内是否存在与所述数据对应的合适的码值。

26. 根据权利要求 25 所述的设备，其特征在于，还包括:

第二存储单元，用于在所述码值区间内不存在与所述数据对应的合适的

码值时，存储所述数据。

27. 一种数据处理的设备，其特征在于，包括:

编码单元，用于采用算术编码算法对数据进行编码，得到码值区间;

获取单元，用于根据所述码值区间获取所述码值;

存储单元，用于存储所述码值;

应用单元，用于根据所述码值进行所述数据的应用操作，所述应用操作包括等值比较、排列和模糊查询中的至少一种。

28. 根据权利要求 27 所述的设备，其特征在于，所述数据为 ID 类字符串，所述应用操作包括等值比较，所述应用单元在所述码值与待比较的码值相等时，确定所述数据与所述待比较的码值对应的数据为相同数据。

29. 根据权利要求 27 所述的设备，其特征在于，所述数据为 ID 类字符串或为字段的字母字符串，所述应用操作包括排列，所述应用单元根据所述码值的大小，确定所述码值在待排列码值中的位置，所述码值的位置用于表示所述数据在与所述待排列的码值对应的数据中的位置。

30. 根据权利要求 27 所述的设备，其特征在于，所述应用操作包括模糊查询，所述应用单元根据所述码值是否在所需模糊查询的前缀字符串的编码区间中，确定所述数据是否包括所述前缀字符串，

其中，在所述码值在所需模糊查询的前缀字符串的编码区间中时，所述数据包括所述前缀字符串，

在所述码值不在所需模糊查询的前缀字符串的编码区间中时，所述数据不包括所述前缀字符串。

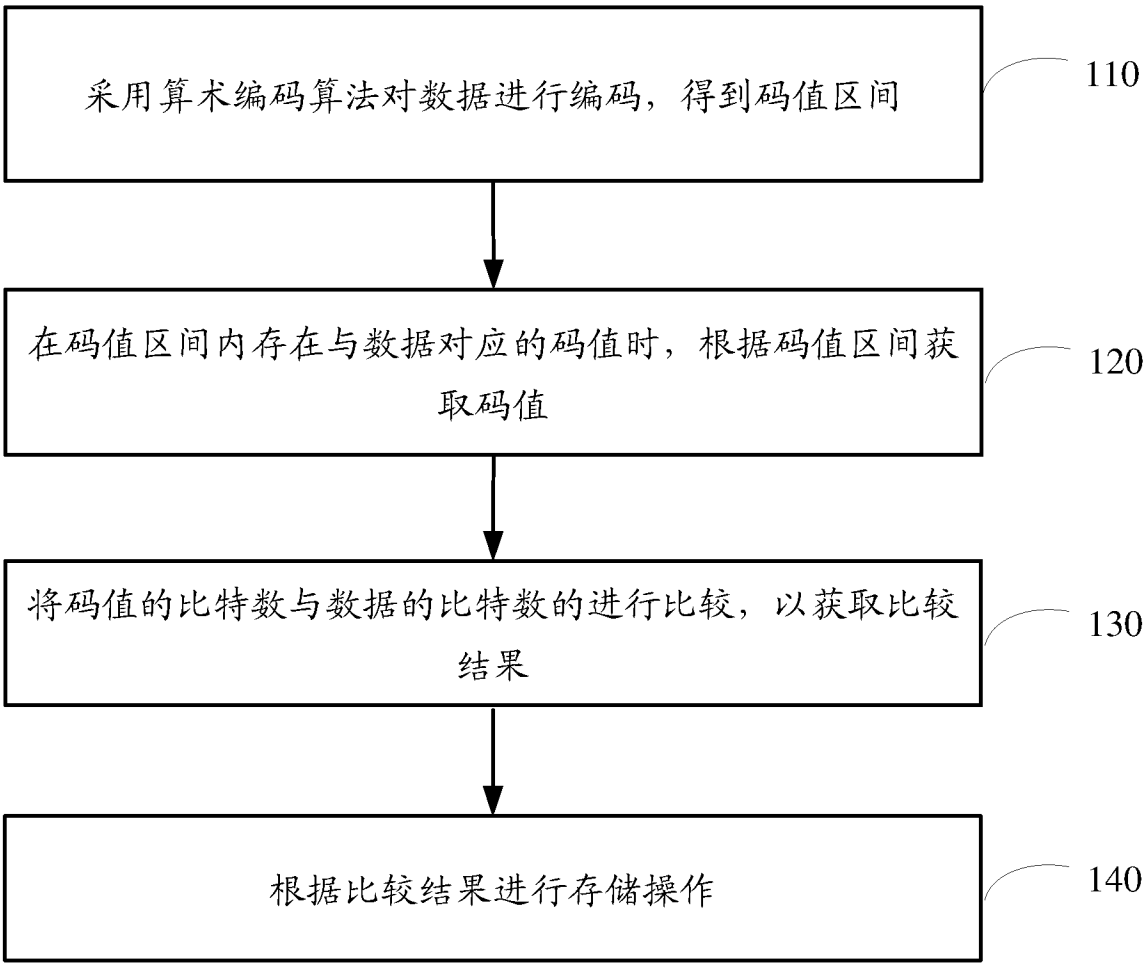

采用算术编码算法对数据进行编码，得到码值区间
110
在码值区间内存在与数据对应的码值时，根据码值区间获取码值
120
将码值的比特数与数据的比特数的进行比较，以获取比较结果
130
根据比较结果进行存储操作
140

图 1

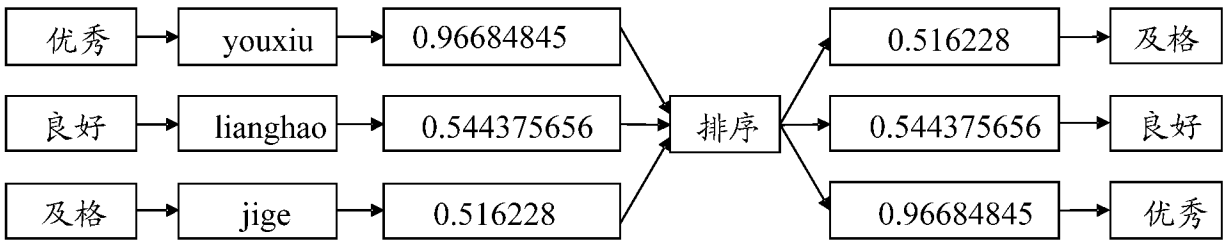

优秀
youxiu
0.96684845
良好
lianghao
0.544375656
及格
jige
0.516228
排序
0.516228
及格
0.544375656
良好
0.96684845
优秀

图 2

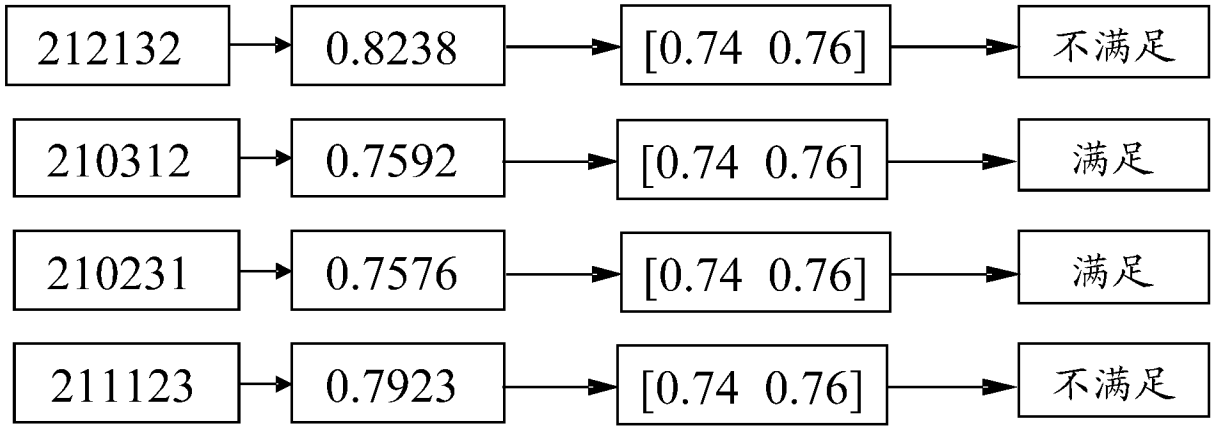

212132
0.8238
[0.74 0.76]
不满足
210312
0.7592
[0.74 0.76]
满足
210231
0.7576
[0.74 0.76]
满足
211123
0.7923
[0.74 0.76]
不满足

图 3

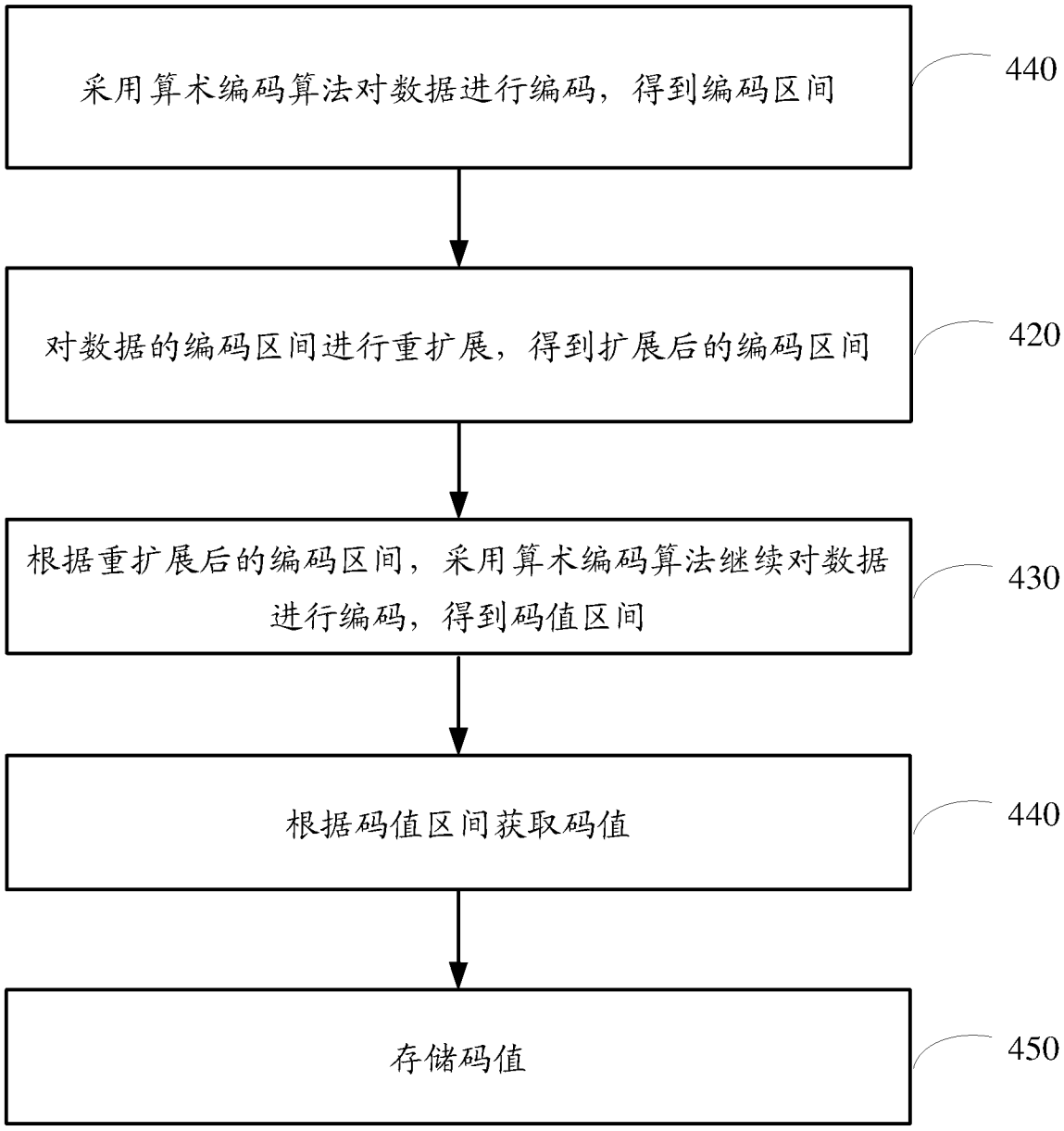

采用算术编码算法对数据进行编码，得到编码区间
440
对数据的编码区间进行重扩展，得到扩展后的编码区间
420
根据重扩展后的编码区间，采用算术编码算法继续对数据进行编码，得到码值区间
430
根据码值区间获取码值
440
存储码值
450

图 4

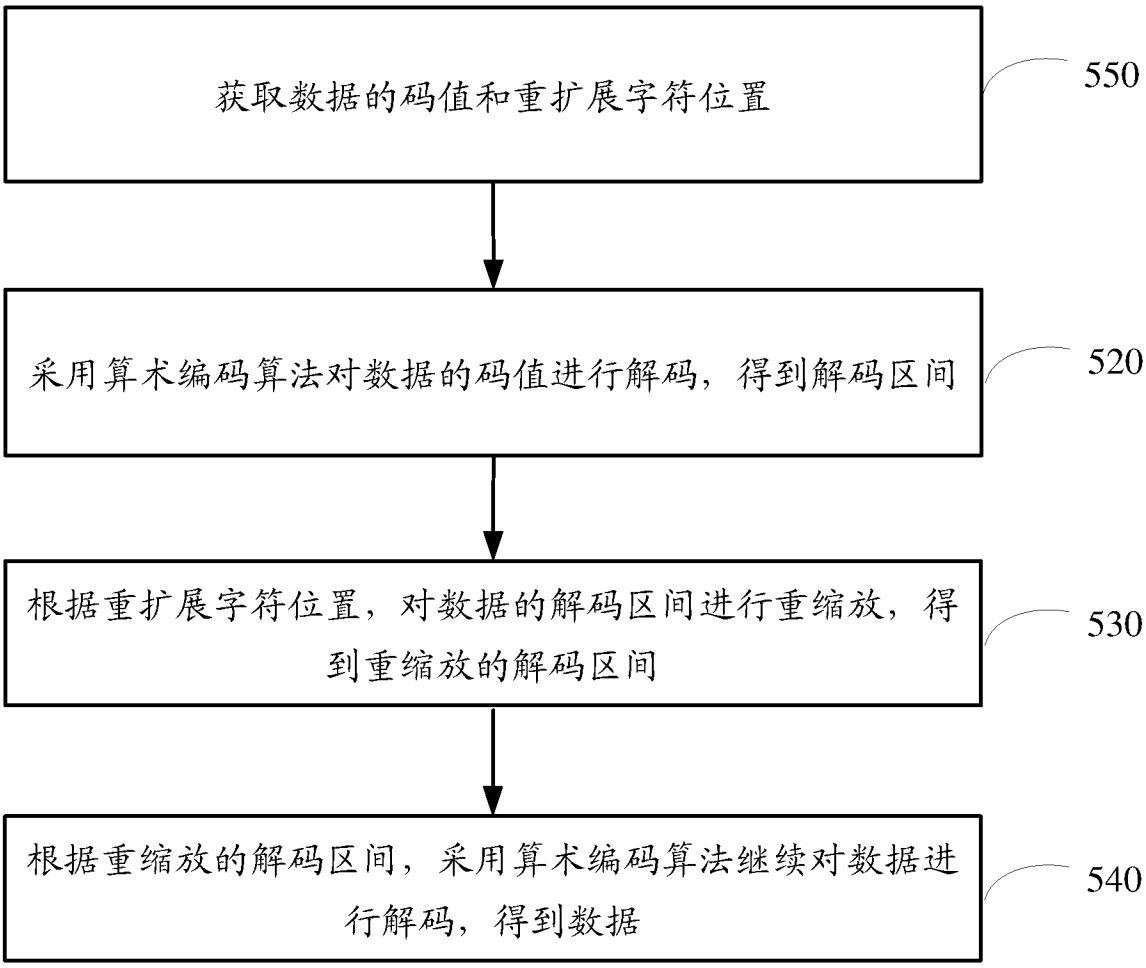

获取数据的码值和重扩展字符位置
550
采用算术编码算法对数据的码值进行解码，得到解码区间
520
根据重扩展字符位置，对数据的解码区间进行重缩放，得到重缩放的解码区间
530
根据重缩放的解码区间，采用算术编码算法继续对数据进行解码，得到数据
540

图 5

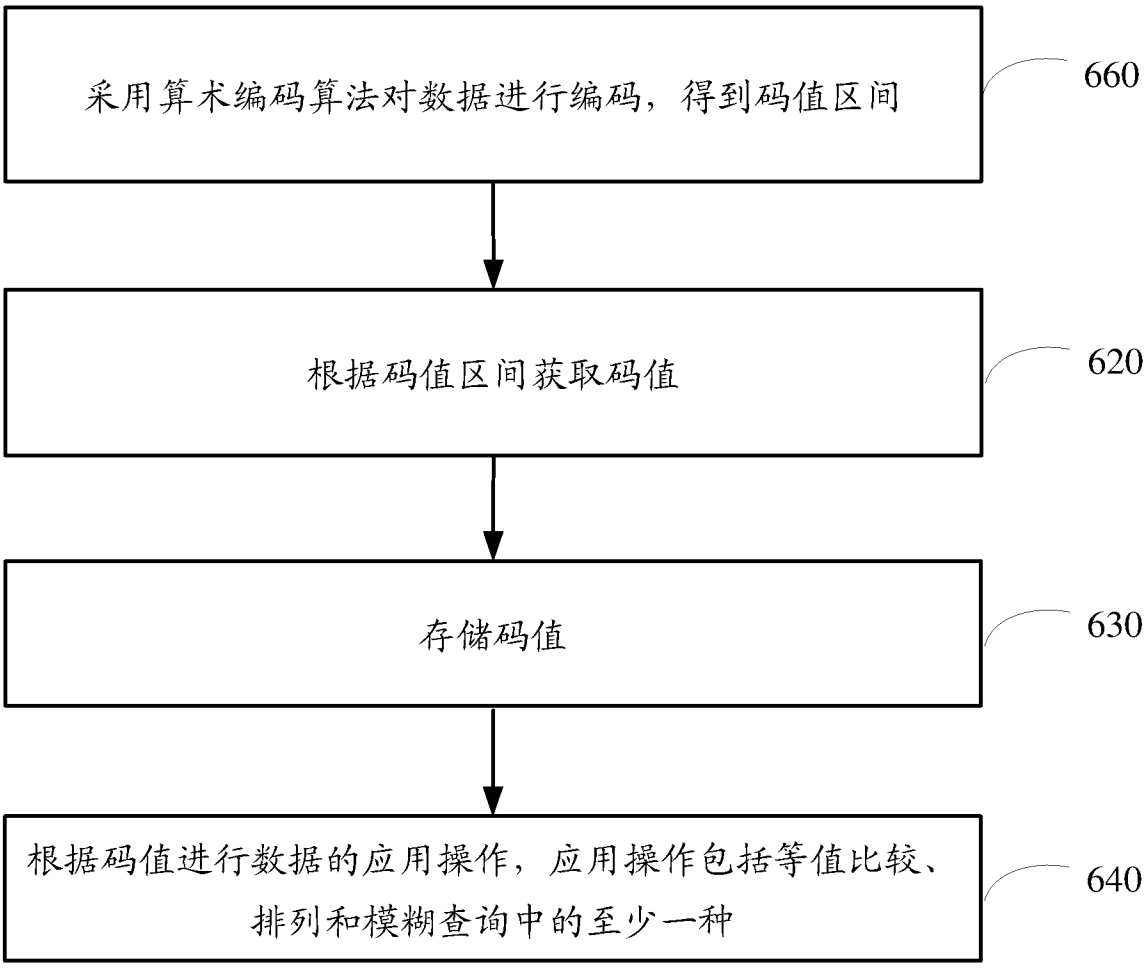

采用算术编码算法对数据进行编码，得到码值区间
660
根据码值区间获取码值
620
存储码值
630
根据码值进行数据的应用操作，应用操作包括等值比较、排列和模糊查询中的至少一种
640

图 6

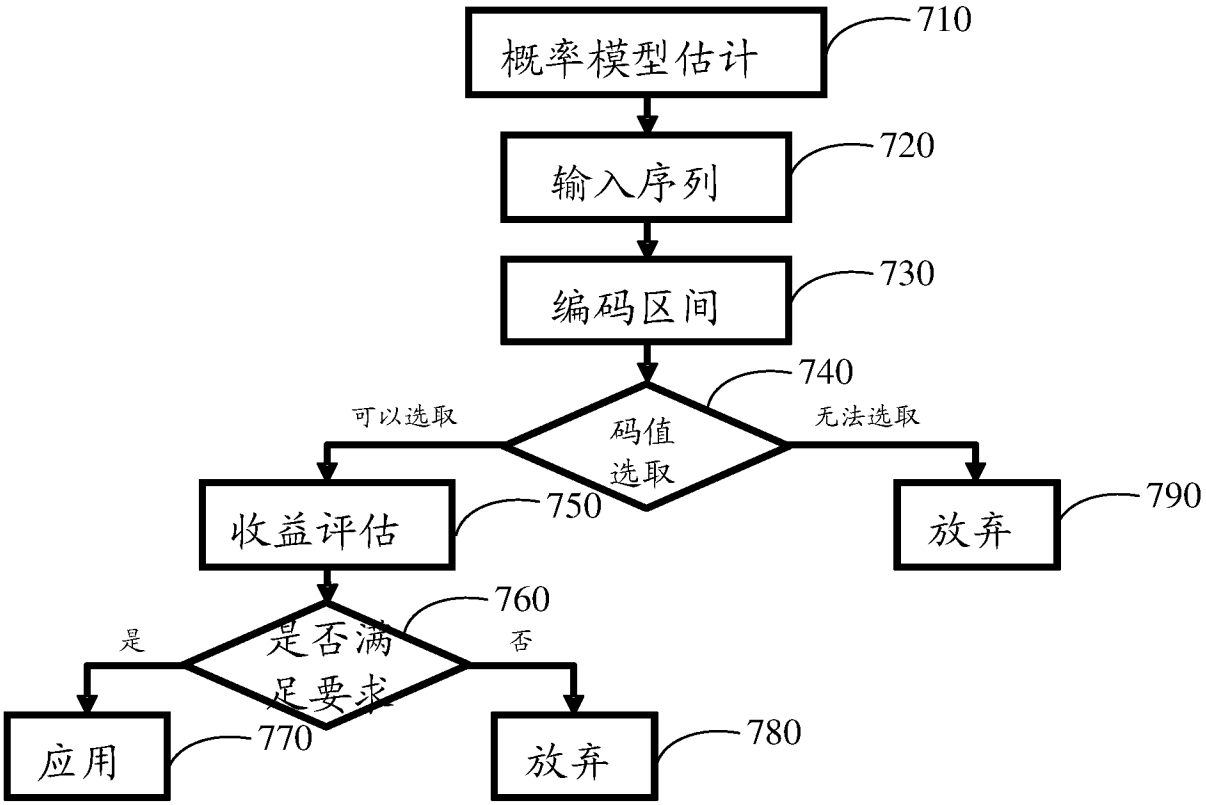

概率模型估计
710
输入序列
720
编码区间
730
码值选取
740
可以选取
无法选取
收益评估
750
放弃
790
是否满足要求
760
是
否
应用
770
放弃
780

图 7

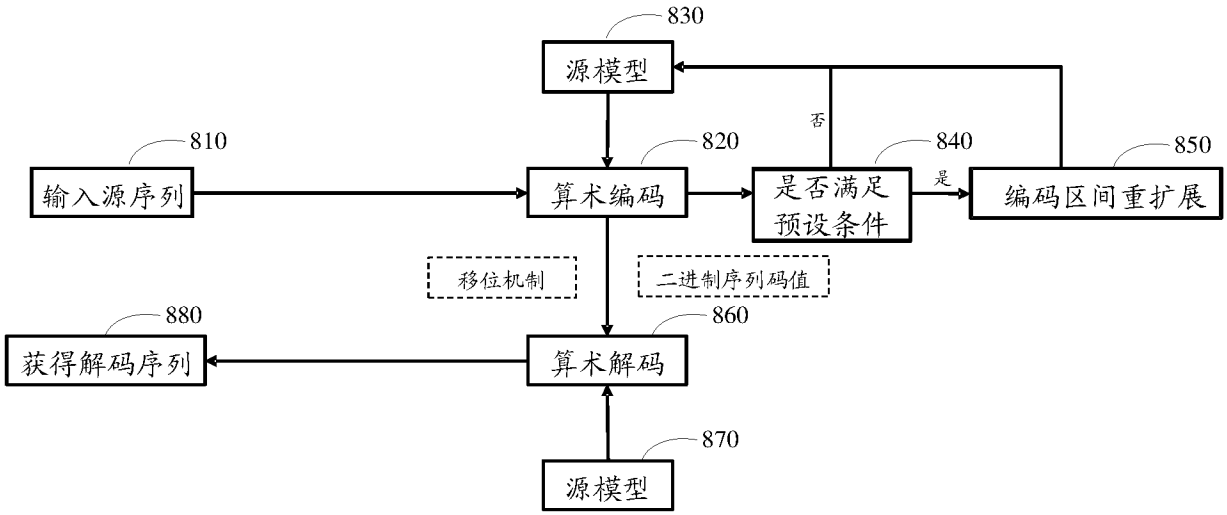

源模型
830
输入源序列
810
算术编码
820
是否满足预设条件
840
否
是
编码区间重扩展
850
移位机制
二进制序列码值
获得解码序列
880
算术解码
860
源模型
870

图 8

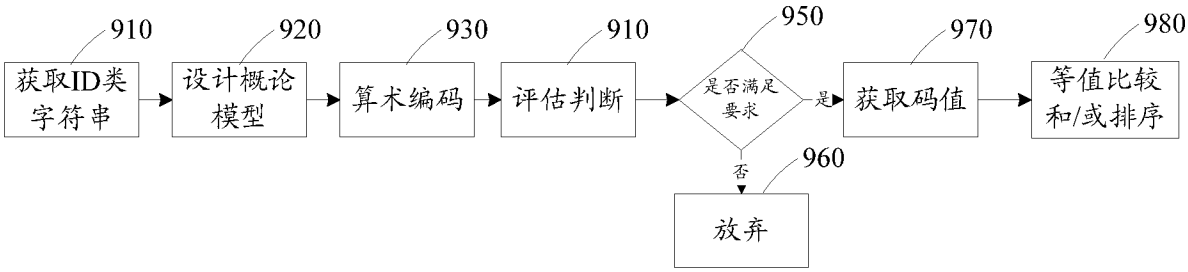

910
获取ID类字符串
920
设计概论模型
930
算术编码
910
评估判断
950
是否满足要求
是
970
获取码值
980
等值比较和/或排序
否
960
放弃

图 9

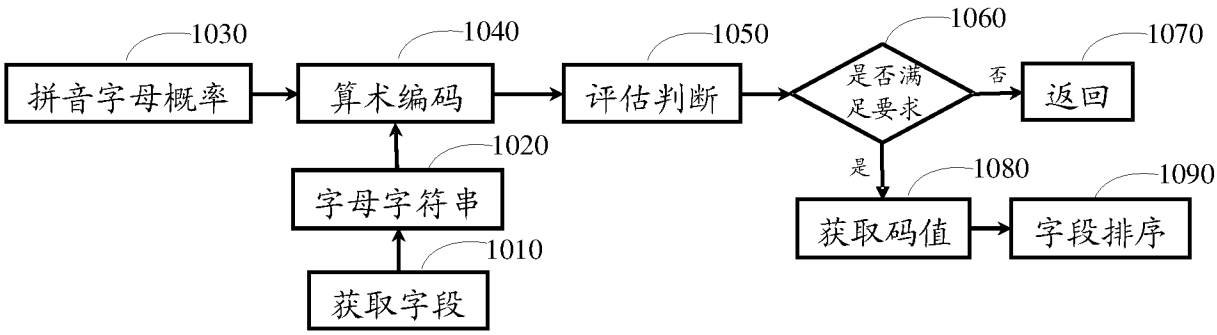

1030
拼音字母概率
1040
算术编码
1050
评估判断
1060
是否满足要求
否
1070
返回
1020
字母字符串
1010
获取字段
是
1080
获取码值
1090
字段排序

图 10

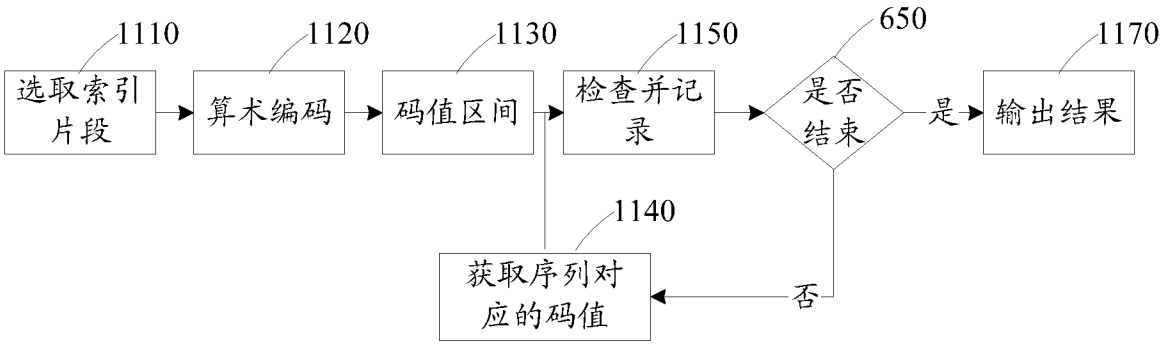

1110
选取索引片段
1120
算术编码
1130
码值区间
1150
检查并记录
650
是否结束
是
1170
输出结果
1140
获取序列对应的码值
否

图 11

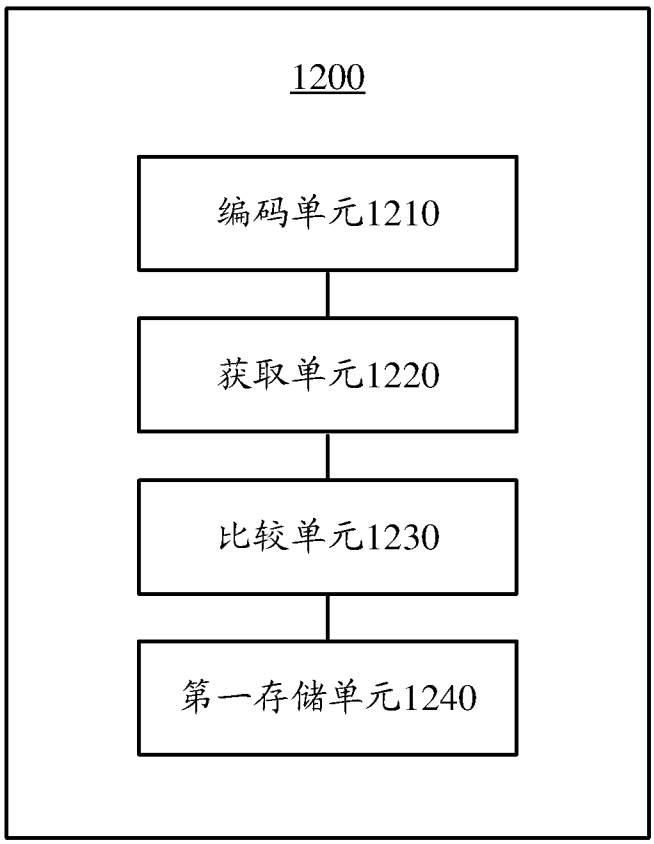

1200
编码单元1210
获取单元1220
比较单元1230
第一存储单元1240

图 12

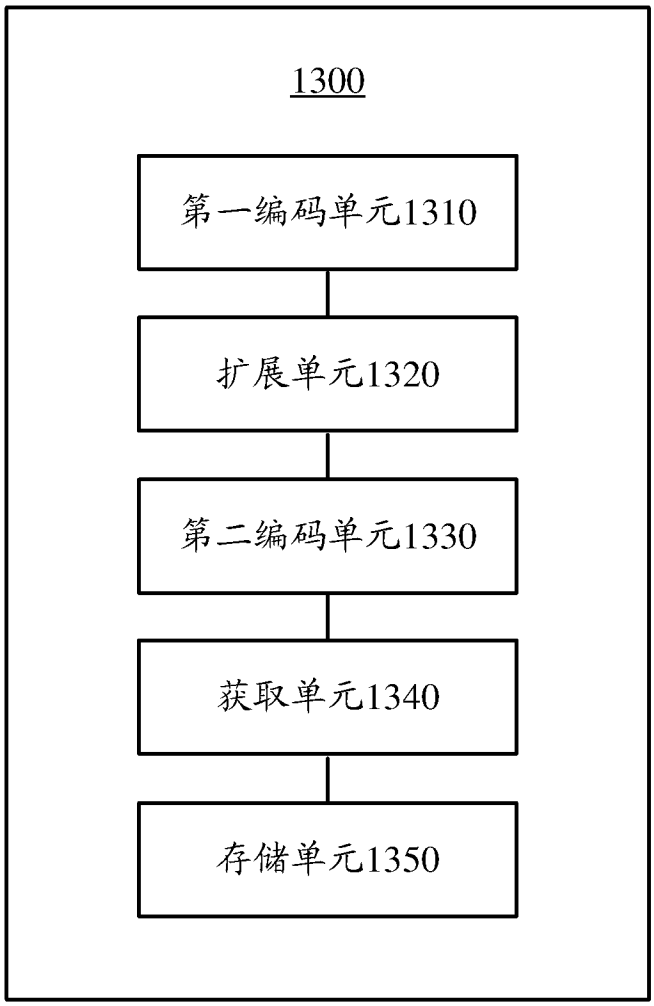


图 13

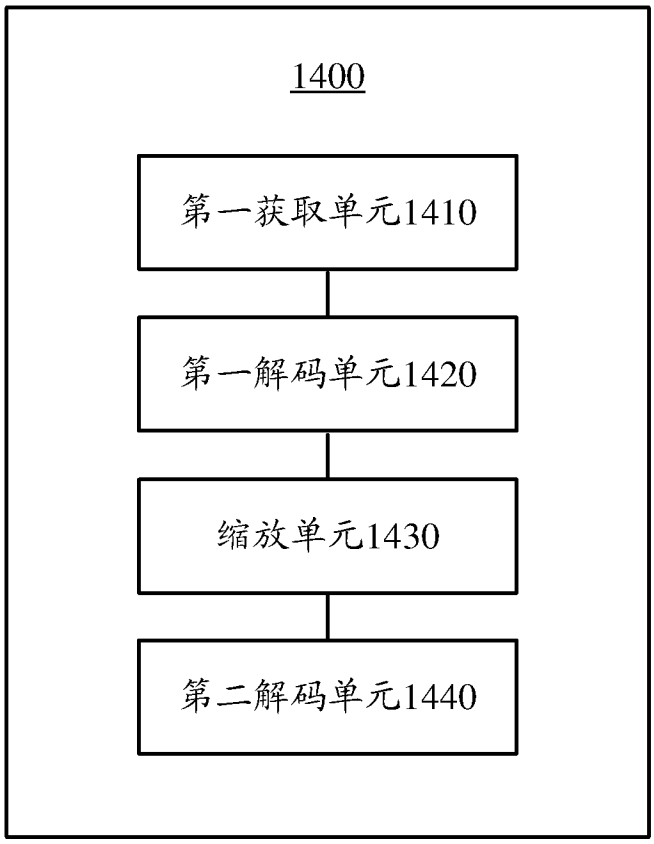

1400
第一获取单元1410
第一解码单元1420
缩放单元1430
第二解码单元1440

图 14

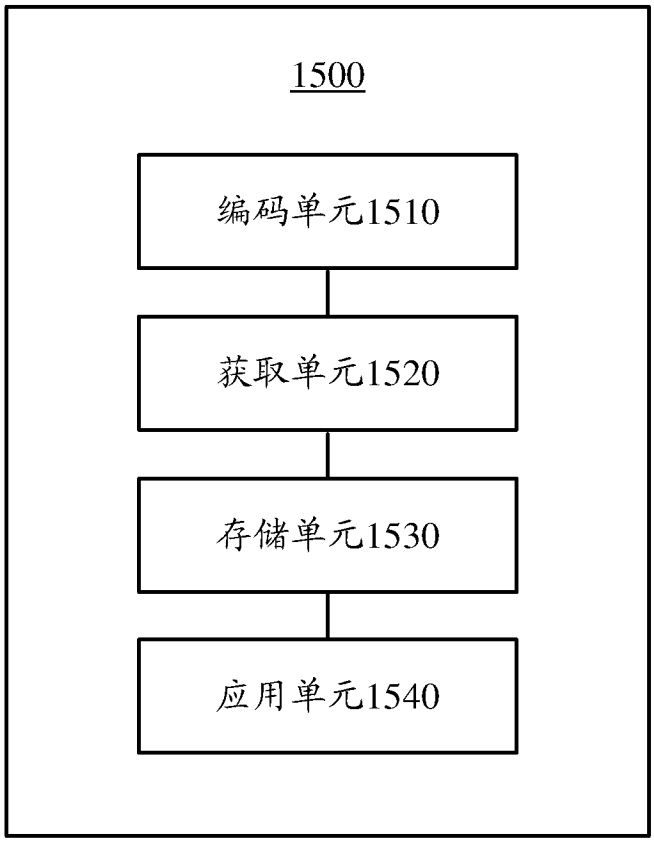

1500
编码单元1510
获取单元1520
存储单元1530
应用单元1540

图 15

1600

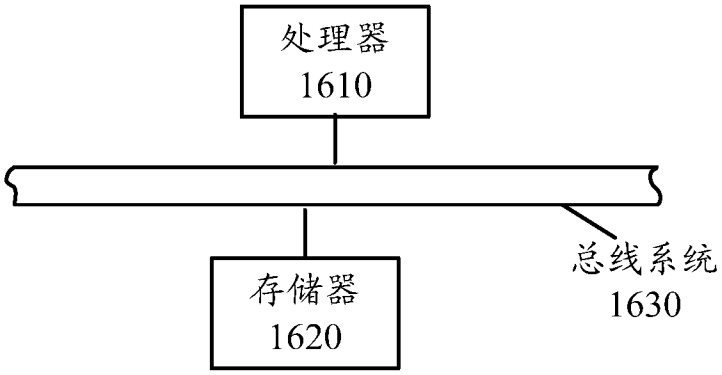


图 16

1700

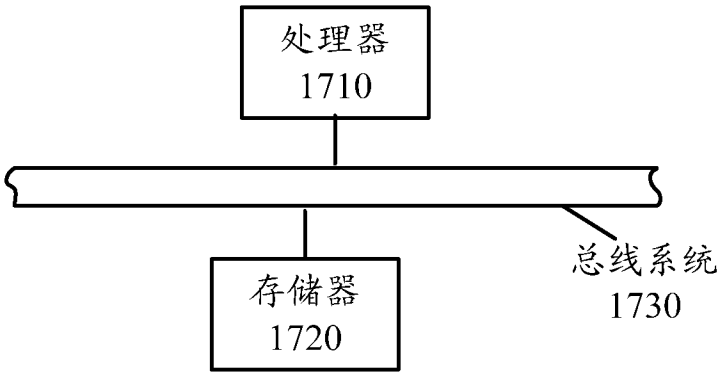


图 17

1800

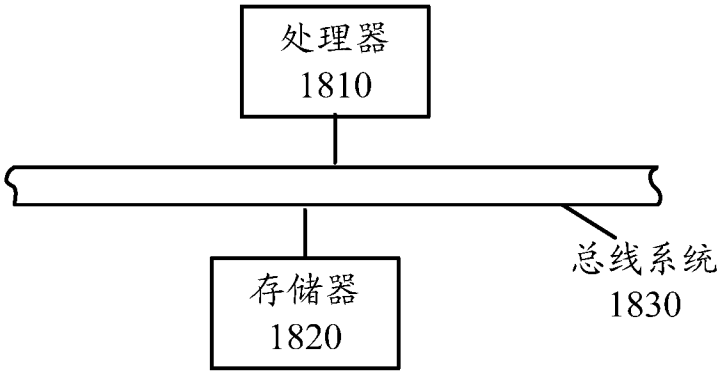


图 18

1900

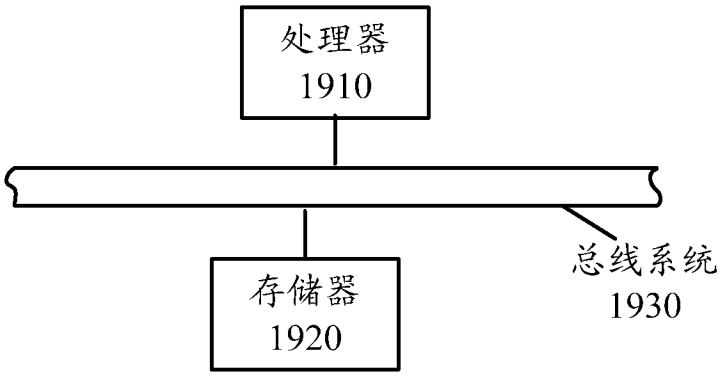


图 19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/070805

A. CLASSIFICATION OF SUBJECT MATTER

H03M 7/30 (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)

H03M; H04N; G06T; 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, CNPAT, CNKI: code value CABAC context based adaptive binary arithmetic coding threshold bit area store memory data equivalence compar+ arrange fuzzy query

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104579360 A (HUAWEI TECHNOLOGIES CO., LTD.), 29 April 2015 (29.04.2015), claims 1-30	1-30
X	CN 101087417 A (PANASONIC CORPORATION), 12 December 2007 (12.12.2007), description, page 2, paragraph 2	1-30
X	CN 1695168 A (SONY CORPORATION), 09 November 2005 (09.11.2005), description, page 10, line 16 to page 17, line 13	1-30

☐ 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
16 March 2016 (16.03.2016)	**12 April 2016 (12.04.2016)**
Name and mailing address of the ISA/CN: 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 **LI, Tingting** Telephone No.: (86-10) **61648276**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/070805

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104579360 A	29 April 2015	None	
CN 101087417 A	12 December 2007	US 2007280353 A1	06 December 2007
		JP 2007329551 A	20 December 2007
CN 1695168 A	09 November 2005	WO 2004034330 A1	22 April 2004
		JP 2004135251 A	30 April 2004
		AU 2003273187 A1	04 May 2004
		KR 20050057642 A1	16 June 2005
		EP 1550081 A1	06 July 2005
		US 2005249289 A1	10 November 2005
		CN 101031086 A	05 September 2007
		KR 20100066586 A	17 June 2010
		KR 20100066585 A	17 June 2010
		US 2010329351 A1	30 December 2010
		US 2010329331 A1	30 December 2010
		US 2010329330 A1	30 December 2010
		KR 20110003542 A	12 January 2011
		US 2011007818 A1	13 January 2011
		SG 167660 A1	28 January 2011
		US 2011044387 A1	24 February 2011
		KR 20110044298 A	28 April 2011
		KR 20110048555 A	11 May 2011
		KR 20110096183 A	29 August 2011
		KR 20110096184 A	29 August 2011
		KR 20110096182 A	29 August 2011
		KR 20110099062 A	05 September 2011
		US 2011216825 A1	08 September 2011
		US 2011216824 A1	08 September 2011
		US 2011216823 A1	08 September 2011
		US 2011216822 A1	08 September 2011
		US 2011222599 A1	15 September 2011
		SG 180037 A1	30 May 2012
		EP 2533203 A2	12 December 2012
		EP 2533204 A2	12 December 2012
		EP 2533207 A2	12 December 2012
		EP 2533206 A2	12 December 2012
		EP 2533205 A2	12 December 2012
		EP 2533208 A2	12 December 2012
		EP 2533209 A2	12 December 2012
		EP 2533210 A2	12 December 2012
		EP 2535870 A2	19 December 2012
		EP 2535871 A2	19 December 2012
		US 2013279579 A1	24 October 2013
		US 2016044313 A1	11 February 2016
		JP 2008042942 A	21 February 2008
		JP 2011061877 A	24 March 2011
		JP 2011061878 A	24 March 2011
		JP 2011087326 A	28 April 2011
		JP 2011101404 A	19 May 2011
		JP 2012178891 A	13 September 2012
		JP 2012178892 A	13 September 2012

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/070805

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
		JP 2012200008 A	18 October 2012
		JP 2012200009 A	18 October 2012
		JP 2012200010 A	18 October 2012
		JP 2012200011 A	18 October 2012
		JP 2012200012 A	18 October 2012
		JP 2014003689 A	09 January 2014
		IN 515DELNP2005 A	19 January 2007

国际检索报告

国际申请号
PCT/CN2016/070805

A. 主题的分类

H03M 7/30(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)

H03M; H04N; G06T; 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,EPODOC,CNPAT,CNKI:算术 编码 码值 区间 数据 比较 阈值 存储 比特 等值 排列 模糊 查询 CABAC context based adaptive binary arithmetic coding threshold bit area store memory data equivalence compar+ arrange fuzzy query

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 104579360 A (华为技术有限公司) 2015年 4月 29日 (2015-04-29) 权利要求1-30	1-30
X	CN 101087417 A (松下电器产业株式会社) 2007年 12月 12日 (2007-12-12) 说明书第2页第2段	1-30
X	CN 1695168 A (索尼株式会社) 2005年 11月 9日 (2005-11-09) 说明书第10页第16行至第17页第13行	1-30

☐ 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 3月 16日	2016年 4月 12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 李婷婷 电话号码 (86-10)61648276

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/070805

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	104579360	A	2015年 4月 29日	无			
CN	101087417	A	2007年 12月 12日	US	2007280353	A1	2007年 12月 6日
				JP	2007329551	A	2007年 12月 20日
CN	1695168	A	2005年 11月 9日	WO	2004034330	A1	2004年 4月 22日
				JP	2004135251	A	2004年 4月 30日
				AU	2003273187	A1	2004年 5月 4日
				KR	20050057642	A1	2005年 6月 16日
				EP	1550081	A1	2005年 7月 6日
				US	2005249289	A1	2005年 11月 10日
				CN	101031086	A	2007年 9月 5日
				KR	20100066586	A	2010年 6月 17日
				KR	20100066585	A	2010年 6月 17日
				US	2010329351	A1	2010年 12月 30日
				US	2010329331	A1	2010年 12月 30日
				US	2010329330	A1	2010年 12月 30日
				KR	20110003542	A	2011年 1月 12日
				US	2011007818	A1	2011年 1月 13日
				SG	167660	A1	2011年 1月 28日
				US	2011044387	A1	2011年 2月 24日
				KR	20110044298	A	2011年 4月 28日
				KR	20110048555	A	2011年 5月 11日
				KR	20110096183	A	2011年 8月 29日
				KR	20110096184	A	2011年 8月 29日
				KR	20110096182	A	2011年 8月 29日
				KR	20110099062	A	2011年 9月 5日
				US	2011216825	A1	2011年 9月 8日
				US	2011216824	A1	2011年 9月 8日
				US	2011216823	A1	2011年 9月 8日
				US	2011216822	A1	2011年 9月 8日
				US	2011222599	A1	2011年 9月 15日
				SG	180037	A1	2012年 5月 30日
				EP	2533203	A2	2012年 12月 12日
				EP	2533204	A2	2012年 12月 12日
				EP	2533207	A2	2012年 12月 12日
				EP	2533206	A2	2012年 12月 12日
				EP	2533205	A2	2012年 12月 12日
				EP	2533208	A2	2012年 12月 12日
				EP	2533209	A2	2012年 12月 12日
				EP	2533210	A2	2012年 12月 12日
				EP	2535870	A2	2012年 12月 19日
				EP	2535871	A2	2012年 12月 19日
				US	2013279579	A1	2013年 10月 24日
				US	2016044313	A1	2016年 2月 11日
				JP	2008042942	A	2008年 2月 21日
				JP	2011061877	A	2011年 3月 24日
				JP	2011061878	A	2011年 3月 24日
				JP	2011087326	A	2011年 4月 28日
				JP	2011101404	A	2011年 5月 19日
				JP	2012178891	A	2012年 9月 13日
				JP	2012178892	A	2012年 9月 13日

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/070805

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)
		JP	2012200008	A	2012年 10月 18日
		JP	2012200009	A	2012年 10月 18日
		JP	2012200010	A	2012年 10月 18日
		JP	2012200011	A	2012年 10月 18日
		JP	2012200012	A	2012年 10月 18日
		JP	2014003689	A	2014年 1月 9日
		IN	515DELNP2005	A	2007年 1月 19日