

# (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 11 月 26 日 (26.11.2020)



(10) 国际公布号  
**WO 2020/232896 A1**

(51) 国际专利分类号:  
**G06N 20/00** (2019.01)

(21) 国际申请号: PCT/CN2019/103434

(22) 国际申请日: 2019 年 8 月 29 日 (29.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:  
201910433025.3 2019 年 5 月 23 日 (23.05.2019) 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: 曹绪文(CAO, Xuwen); 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: 深圳市隆天联鼎知识产权代理有限公司(SHENZHEN LUNGTIN LIANDING INTELLECTUAL PROPERTY AGENT LTD.); 中国广东省深圳市福田区南园路上田大厦 4A 刘抗美, Guangdong 518000 (CN)。

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

(54) Title: DATA DICTIONARY GENERATION METHOD AND APPARATUS, ELECTRONIC DEVICE AND STORAGE MEDIUM

(54) 发明名称: 数据字典生成的方法、装置、电子设备和存储介质

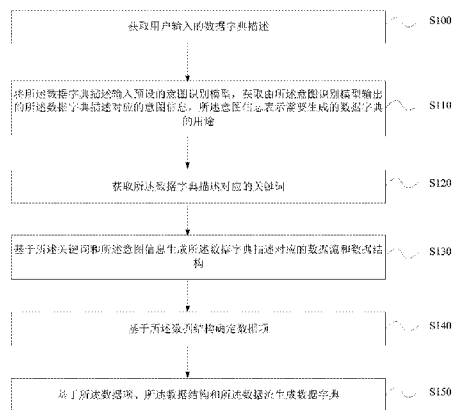


图 1

S100 Acquire a data dictionary description inputted by a user  
S110 Input the data dictionary description into a preset intent recognition model, and acquire intent information which is outputted by the intent recognition model and corresponds to the data dictionary description, the intent information representing the use of a data dictionary to be generated  
S120 Acquire a keyword corresponding to the data dictionary description  
S130 Generate, on the basis of the keyword and the intent information, a data stream and a data structure corresponding to the data dictionary description  
S140 Determine a data item on the basis of the data structure  
S150 Generate a data dictionary on the basis of the data item, the data structure, and the data stream

(57) Abstract: Disclosed are a data dictionary generation method and apparatus, an electronic device and a storage medium, relating to the field of data processing. Said method comprises: acquiring a data dictionary description inputted by a user; inputting the data dictionary description into a preset intent recognition model, and acquiring intent information which is outputted by the intent recognition model and corresponds to the data dictionary description, the intent information representing the use of a data dictionary to be generated; acquiring a keyword corresponding to the data dictionary description; generating, on the basis of the keyword and the intent information, a data stream and a data structure corresponding to the data dictionary description; determining a data item on the basis of the data structure; and generating a data dictionary on the basis of the data item, the data structure, and the data stream. The technical solution of the embodiments improves the efficiency of generating a data dictionary.

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))。

---

(57) 摘要: 公开了一种数据字典生成的方法、装置、电子设备和存储介质, 涉及数据处理领域, 该方法包括: 获取用户输入的数据字典描述; 将所述数据字典描述输入预设的意图识别模型, 获取由所述意图识别模型输出的所述数据字典描述对应的意图信息, 所述意图信息表示需要生成的数据字典的用途; 获取所述数据字典描述对应的关键词; 基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构; 基于所述数据结构确定数据项; 基于所述数据项、所述数据结构和所述数据流生成数据字典。实施例的技术方案提高了数据字典生成的效率。

## 数据字典生成的方法、装置、电子设备和存储介质

本申请要求 2019 年 5 月 23 日递交、发明名称为“数据字典生成的方法、装置、电子设备和存储介质”的中国专利申请 201910433025.3 的优先权，在此通过引用将其全部内容合并于此。

### 技术领域

本发明涉及数据处理领域，特别是涉及数据字典生成的方法、装置、电子设备和存储介质。

### 背景技术

数据字典是指对数据的数据项、数据结构、数据流、数据存储、处理逻辑等进行定义和描述，对系统中使用的所有数据元素的定义的集合。

本申请的发明人意识到，在开发软件系统的过程中，程序工作人员通常要根据依据要生成的数据字典的注释信息来生成不止一个数据字典来确保要开发的软件系统中数据的一致性，从而在制作数据字典上花费大量的时间。

因此，如何能快速生成满足需要的数据字典，节省程序工作人员的时间，缩短软件系统的开发时间是亟待解决的问题。

### 发明内容

基于此，本申请实施例提供了一种数据字典生成的方法、装置、电子设备和存储介质，以至少解决生成数据字典效率低的问题。

根据本申请实施例的第一方面，提供了一种数据字典生成的方法，包括：获取用户输入的数据字典描述；将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；获取所述数据字典描述对应的关键词；基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；基于所述数据结构确定数据项；基于所述数据项、所述数据结构和所述数据流生成数据字典。

根据本申请的第二方面方面，提供了一种数据字典生成的装置，包括：第一获取模块，用于获取用户输入的数据字典描述；第二获取模块，用于将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；第三获取模块，获取所述数据字典描述对应的关键词；第一生成模块：根据所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；确定模块，用于基于所述数据结构确定数据项；第二生成模块，用于基于所述数据项、所述数据结构和所述数据流生成数据字典。

根据本申请的第三方面方面，提供了一种数据字典生成的电子设备，包括：存储器，配置为存储可执行指令；处理器，配置为执行存储器中存储的可执行指令，以执行以上所述的方法。

根据本申请的第四方面，提供了一种计算机非易失性可读存储介质，其存储有计算机程序指令，当所述计算机指令被计算机执行时，使计算机执行以上所述的方法。

在本申请实施例所提供的技术方案中，通过获取用户端输入的数据字典描述，获取所述数据字典描述的意图信息和关键词，基于所述关键词和所述意图信息获取所述数据字典描述对应的数据流和数据结构，基于所述数据结构包含的数据项名确定数据项，再依据所述数据流、数据结构和数据项生成数据字典，使得开发人员在需要生成数据字典时，只需提供数据字典描述即可，进而本发明实施例的技术方案可以根据开发人员提供的数据字典描述自动生成数据字典，解决了开发人员手动制作数据字典而需要花费较多时间的问题，提高了数据字典生成的效率。

本发明的其他特性和优点将通过下面的详细描述变得显然，或部分地通过本发明的实践而习得。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性的，并不能限制本发明。

### 附图说明

图 1 示出根据本发明示例实施方式的数据字典生成的流程图。

图 2 示出根据本发明一示例实施方式的在将所述数据字典描述输入预设的意图识别模型之前的流程图。

图 3 示出根据本发明一示例实施方式的获取所述数据字典描述对应的关键词的详细流程图。

图 4 示出根据本发明一示例实施方式的获取所述数据字典描述对应的关键词的详细流程图。

图 5 示出根据本发明一示例实施方式的基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构的详细流程图。

图 6 示出根据本发明一示例实施方式的基于所述数据结构确定数据项的详细流程图。

图 7 示出根据本发明一示例实施方式的基于所述数据项、所述数据结构和所述数据流生成数据字典之后的流程图。

图 8 示出根据本发明一示例实施方式的数据字典生成的装置。

图 9 示出根据本发明一示例实施方式的数据字典生成的系统架构图。

图 10 示出根据本发明一示例实施方式的数据字典生成的电子设备的框图。

图 11 示出根据本发明一示例实施方式的数据字典生成的计算机非易失性可读存储介质的示意图。

### 具体实施方式

现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的范例；相反，提供这些实施方式使得本发明将更加全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中，提供许多具体细节从而给出对本发明的实施方式的充分理解。然而，本领域技术人员将意识到，可以实践本发明的技术方案而省略所述特定细节中的一个或更多，或者可以采用其它的方

法、组元、装置、步骤等。在其它情况下，不详细示出或描述公知技术方案以避免喧宾夺主而使得本发明的各方面变得模糊。

此外，附图仅为本发明的示意性图解，并非一定是按比例绘制。图中相同的附图标记表示相同或类似的部分，因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体，不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体，或在一个或多个硬件模块或集成电路中实现这些功能实体，或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

图 1 示出根据本发明一示例实施方式的数据字典生成的流程图，可以包括如下步骤：

步骤 S100：获取用户输入的数据字典描述；

步骤 S110：将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；

步骤 S120：获取所述数据字典描述对应的关键词；

步骤 S130：基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；

步骤 S140：基于所述数据结构确定数据项；

步骤 S150：基于所述数据项、所述数据结构和所述数据流生成数据字典。

下面，将结合附图对本示例实施方式中上述数据字典生成的各步骤进行详细的解释以及说明。

在步骤 S100 中：获取用户输入的数据字典描述。

在本发明的一个实施例中，数据字典描述是指对数据字典的用途的一段说明文本。可通过用户端获取所述数据字典描述，也可通过获取用户输入的所述数据字典描述对应的语音信息后，基于语音识别模型来获取所述数据字典描述，进而基于所述数据字典描述，使得服务器可以对所述数据字典描述进行数据分析，从而生成所述数据字典描述对应的数据字典。

在一实施例中，服务器通过用户端获取用户输入的数据字典描述。

在步骤 S110 中：将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途。

通过意图识别模型，使得服务器可以快速获取所述用户输入的数据字典描述对应的需要生产的数据字典的用途，便于对数据字典描述进行分析。

在一实施例中，用户输入的数据字典描述是“用于记录学生选课情况，尤其要记录选课学生的学号、年龄、性别”，将该数据字典描述输入意图识别模型，得到该数据字典描述对应的意图信息为“记录学生选课信息”。

在一实施例中，如图 2 所示，在步骤 S110 之前还包括：

步骤 S107：获取预先设置的数据字典描述样本集合；

步骤 S108：识别所述数据字典描述样本集合中的数据字典描述样本对应的意图信息；

步骤 S109：将所述数据字典描述样本输入所述意图识别模型，由所述意图识别模型

输出意图信息,将所述意图识别模型输出的意图信息和识别出的所述数据字典描述样本对应的意图信息进行比对,如不一致,则调整所述意图识别模型的参数,直至所述意图识别模型输出的意图信息与识别出的所述数据字典描述样本对应的意图信息一致。

5 本发明上述实施例的技术方案通过意图识别模型的方式获取所述数据字典描述对应的意图信息,不仅处理速度更快而且判断标准统一,避免了因对意图信息判断标准不统一造成输出的数据字典描述对应的意图信息不一致而导致生成的数据字典出现错误的情形。

继续参照图 1 所示,在步骤 S120 中:获取所述数据字典描述对应的关键词。

10 关键词是表明文本信息中重要或主要信息的词汇,通过获取所述数据字典描述对应的关键词,进而对所述关键词进行分析,获取所述数据字典描述对应的数据结构和数据流,从而生成所述数据字典描述对应的数据字典。

在本发明的一个实施例中,步骤 S120 中获取数据字典描述对应的关键词可以有多种实现方式,以下介绍其中的两种实现方式:

实现方式一:

在一实施例中,如图 3 所示,步骤 S120 包括:

15 步骤 S1201:将所述数据字典描述分句;

步骤 S1202:将所述数据字典描述分成的句子与预设的数据字典描述句模板库中的数据字典描述句模板进行比对,以确定与所述数据字典描述分成的句子相匹配的数据字典描述句模板;

20 步骤 S1203:按照所述数据字典描述句模板中规定的数据字典描述关键词的位置,确定所述分成的句子中的关键词。

图 3 所示实施例的技术方案通过对所述数据字典描述分句,基于所述数据字典描述分句确定与之对应的预存的数据字典描述句模板,再基于所述数据字典描述句模板中标明的关键词位置确定所述数据字典描述分句的关键词,可以准确快速的提取数据字典描述对应的关键词。

25 在一实施例中,可以基于段落间的标点符号对数据字典描述进行分句,如将句号或逗号间的文本信息作为一个句子。

在一实施例中,假设数据字典描述是“用于记录学生选课情况,尤其要记录选课学生的学号、年龄及性别”,那么对该数据字典描述分句得到“用于记录学生选课情况”,“尤其要记录选课学生的学号、年龄及性别”,确定的该数据字典描述分句对应的数据字典描述句模板为“用于记录\_\_情况”“尤其要记录\_\_的\_\_及\_\_”,其中下划线部分为标明的关键词位置,通过该数据字典描述分句对应的数据字典描述句模板中标明的关键词位置,确定该数据字典描述对应的关键词为“学生选课”“选课学生”“学号”“年龄”“性别”。

30

实现方式二:

在一实施例中,如图 4 所示,所述步骤 S120 包括:

35 步骤 S1201':获取预先设置的数据字典描述样本集合;

步骤 S1202':确定所述数据字典描述样本集合中的数据字典描述样本的关键词;

步骤 S1203':将所述数据字典描述样本输入第一机器学习模型,由所述第一机器学习模型输出关键词,将所述第一机器学习模型输出的关键词与确定的所述数据字典描述样

本的关键词进行比对,如不一致,则调整所述第一机器学习模型的参数,直至所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词一致;

步骤 S1204': 将所述数据字典描述输入所述第一机器学习模型,获取由所述第一机器学习模型输出所述数据字典描述对应的关键词。

5 图 4 所示实施例的技术方案通过预设机器学习模型的方式,可快速获取所述数据字典描述对应的关键词信息,相对于数据字典描述分句模板的方式获取所述数据字典描述对应的关键词,因通过数据字典描述分句模版的方式中数据字典描述分句模版需要人为预先设定,而不同人对于所述数据字典描述分句模版的设定其标准并不统一,故通过预设机器学习模型的方式获取所述数据字典描述对应的关键词,其标准更为统一。

10 继续参照图 1 所示,在步骤 S130 中:基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构。

在一实施例中,如图 5 所示,步骤 S130 包括:

步骤 S1301: 获取预先设置的数据字典描述对应的关键词和意图信息样本集合;

15 步骤 S1302: 确定所述数据字典描述对应的关键词和意图信息样本对应的数据流和数据结构;

步骤 S1303: 将所述数据字典描述对应的关键词和意图信息样本输入所述第二机器学习模型,由所述第二机器学习模型输出数据流与数据结构,将所述第二机器学习模型输出的数据流与数据结构与确定的数据流与数据结构进行比对,如不一致,则调整所述第二机器学习模型的参数,直至所述第二机器学习模型输出的数据流和数据结构与确定的数据流与数据结构一致;

20 步骤 S1304: 将所述数据流和数据结构输入所述第二机器学习模型,获取由所述第二机器学习模型输出所述数据流和数据结构。

图 5 所示实施例的技术方案,通过第二机器学习模型的方式相对于现有技术中通过人工及调用预设的模板的方式来生成数据流和数据结构而言,可快速的获取所述数据字典描述对应的数据流和数据结构。在一实施例中,用户输入的数据字典描述对应的关键词为“学生选课”“学号”“性别”“年龄”,用户输入的数据字典描述对应的意图信息为“记录学生选课信息”,将用户输入的数据字典描述对应的关键词和意图信息输入第二机器学习模型,得到的数据结构为学生所选的对应课程,其中学生包括:“姓名”“学号”“年龄”“性别”,课程包括“课程编号”“选课编号”,数据流为记录学生选课信息,数据流来源为学生选课处理,数据流去向为学生选课储存,数据流组成为:“学号”“课程编号”。

30 在一实施例中,所述第一机器学习模型与所述第二机器学习模型均为预先训练好的机器学习模型,其中所述第一机器学习模型用于关键词的提取,第二机器学习模型用于基于所述关键词于所述意图信息生成数据流与数据结构。

继续参照图 1 所示,在步骤 S140 中:基于所述数据结构确定数据项。

35 在一实施例中,如图 6 所示,步骤 S140 包括:

步骤 S1401: 基于所述数据结构确定数据项名;

步骤 S1402: 基于所述数据项名在预设数据库中匹配相应的数据项。

图 6 所示实施例的技术方案通过数据结构包含的数据项名与数据项的对应关系,确定所述数据字典描述对应的数据项,以便于服务器依据所述数据项、数据流和数据结构生成

数据字典。

在一实施例中，数据结构可以为：学生所选的对应课程，其中学生包括：“姓名”“学号”“年龄”“性别”，课程包括“课程编号”“选课编号”，提取该数据结构中的数据项名：姓名、年龄、性别、课程编号、选课编号，可确定数据项为：姓名、年龄、性别、课程编号、选课编号。

步骤 S150：基于所述数据项、所述数据结构和所述数据流生成数据字典。

在一实施例中，假设数据项为姓名、年龄、性别、课程编号、选课编号，数据结构为学生所选的对应课程，其中学生包括：“姓名”“学号”“年龄”“性别”，课程包括“课程编号”“选课编号”，数据流为记录学生选课信息，数据流来源为学生选课处理，数据流去向为学生选课储存，数据流组成为：“学号”“选课编号”，那么生成的数据字典可以如表 1 所示：

| 序号 | 表名      |
|----|---------|
| 1  | 学生基本信息表 |
| 2  | 选课信息表   |

表 1

学生基本信息如表 2 所示：

| 名称 | 数据类型      | 主键  | 非空  | 约束条件          |
|----|-----------|-----|-----|---------------|
| 学号 | char (10) | Yes | Yes |               |
| 姓名 | varchar   | No  | Yes |               |
| 性别 | char (2)  | No  | Yes | In “男” or “女” |

表 2

选课信息如表 3 所示：

| 名称   | 数据类型     | 主键  | 非空  | 约束条件 |
|------|----------|-----|-----|------|
| 选课编号 | char (4) | Yes | Yes |      |
| 课程编号 | char (4) | No  | Yes |      |

表 3

在一实施例中，如图 7 所示，在图 1 所示的各个步骤之后，本发明实施例提供的数据字典生成的方法还可以包括如下步骤：

步骤 S160：将所述数据字典存储至共享池，并赋予相应的等级；

步骤 S170：获取用户登录信息；

步骤 S180: 基于所述用户登录信息中包含的该用户的权限信息和该用户所要调取的数据字典的等级, 确定所述用户是否能调取相应的数据字典。

图 7 所示实施例的技术方案通过将所述数据字典存储至共享池的方式, 使得满足预设条件的用户都可以获取所述数据字典, 实现资源共享, 通过判断所述用户登录信息中包含的该用户的权限信息是否满足调取该用户所要调取的数据字典等级的方式, 防止高等级数据字典被不满足调取该数据字典的用户获取并泄露的风险。

在一实施例中, 根据用户的登录信息确定该用户的权限为可调取等级小于等于 5 级的数据字典, 该用户要调取的数据字典为 7 级, 因该用户可调取的等级 5 小于要调取数据字典的等级 7, 故该用户无法从共享池中调取该等级为 7 级的数据字典。

本发明还提供了一种数据字典生成的装置。参考图 8 所示, 所述数据字典生成的装置 800 包括: 第一获取模块 810、第二获取模块 820、第三获取模块 830、第一生成模块 840、确定模块 850、第二生成模块 860。其中:

第一获取模块 810: 用于获取用户输入的数据字典描述;

第二获取模块 820: 用于将所述数据字典描述输入预设的意图识别模型, 获取由所述意图识别模型输出的所述数据字典描述对应的意图信息, 所述意图信息表示需要生成的数据字典的用途;

第三获取模块 830: 用于获取所述数据字典描述对应的关键词;

第一生成模块 840: 用于根据所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构;

确定模块 850: 用于基于所述数据结构确定数据项;

第二生成模块 860, 用于基于所述数据项、所述数据结构和所述数据流生成数据字典。

在一实施例中, 第三获取模块 830 还可配置为: 将所述数据字典描述分句, 将所述数据字典描述分成的句子与预设的数据字典描述句模板库中的数据字典描述句模板进行比对, 以确定与所述数据字典描述分成的句子相匹配的数据字典描述句模板, 按照所述数据字典描述句模板中规定的数据字典描述关键词的位置, 确定所述分成的句子中的关键词。

在一实施例中, 第三获取模块 830 还可配置为: 获取预先设置的数据字典描述样本集合; 确定所述数据字典描述样本集合中的数据字典描述样本的关键词; 将所述数据字典描述样本输入第一机器学习模型, 由所述第一机器学习模型输出关键词, 将所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词进行比对, 如不一致, 则调整所述第一机器学习模型的参数, 直至所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词一致; 将所述数据字典描述输入所述第一机器学习模型, 获取由所述第一机器学习模型输出所述数据字典描述对应的关键词。

在一实施例中, 第一生成模块 840 还可配置为: 获取预先设置的数据字典描述对应的关键词和意图信息样本集合, 确定所述数据字典描述对应的关键词和意图信息样本对应的数据流和数据结构, 将所述数据字典描述对应的关键词和意图信息样本输入所述第二机器学习模型, 由所述第二机器学习模型输出数据流与数据结构, 将所述第二机器学习模型输出的数据流与数据结构与确定的数据流与数据结构进行比对, 如不一致, 则调整所述第二机器学习模型的参数, 直至所述第二机器学习模型输出的数据流与数据结构与确定的数据流与数据结构一致, 将所述数据流和数据结构输入所述第二机器学习模型, 获取由所述第二机器学习模型输出所述数据流和数据结构。

在一实施例中，确定模块 850 还可配置为：基于所述数据结构确定数据项名，基于所述数据项名在预设数据库中匹配相应的数据项。

在一实施例中，所述数据字典生成装置 800 还包括：意图模型训练模块，用于获取预先设置的数据字典描述样本集合，识别所述数据字典描述样本集合中的数据字典描述样本对应的意图信息，将所述数据字典描述样本输入所述意图识别模型，由所述意图识别模型输出意图信息，将所述意图识别模型输出的意图信息和识别出的所述数据字典描述样本对应的意图信息进行比对，如不一致，则调整所述意图识别模型的参数，直至所述意图识别模型输出的意图信息与识别出的所述数据字典描述样本对应的意图信息一致。

在一实施例中，所述数据字典生成装置 800 还包括：共享模块，用于将所述数据字典存储至共享池，并赋予相应的等级，获取用户登录信息，基于所述用户登录信息中包含的该用户的权限信息和该用户所要调取的数据字典的等级，确定所述用户是否能调取相应的数据字典。

上述数据字典生成的装置中各模块的具体细节已经在对应的方法中进行了详细的描述，因此此处不再赘述。

应当注意，尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元，但是这种划分并非强制性的。实际上，根据本发明的实施方式，上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之，上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

此外，尽管在附图中以特定顺序描述了本发明中方法的各个步骤，但是，这并非要求或者暗示必须按照所述特定顺序来执行这些步骤，或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的，可以省略某些步骤，将多个步骤合并为一个步骤执行，以及/或者将一个步骤分解为多个步骤执行等。

通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本发明实施方式的技术方案可以以软件产品的形式体现出来，所述软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、移动终端、或者网络设备等）执行根据本发明实施方式的方法。

图 9 示出根据本发明一示例实施方式的数据字典生成的系统架构框图。所述系统架构包括：用户端 910、服务器 920。

在一实施例中，服务器 920 通过用户端 910 获取用户输入的数据字典描述，服务器 920 根据所述数据字典描述，获取所述数据字典描述对应的意图信息和关键词，服务器 920 根据所述意图信息和所述关键词确定所述数据字典描述对应的数据流和数据结构，服务器 920 根据所述数据结构中数据项名确定数据项，服务器 920 基于所述数据项、数据项结构和数据流生成所述数据字典描述对应的数据字典。

通过以上对系统架构的描述，本领域的技术人员易于理解，这里描述的系统架构能够实现图 8 所示的数据字典生成的装置中各个模块的功能。

在本发明的示例性实施例中，还提供了一种能够实现上述方法的电子设备。

所属技术领域的技术人员能够理解，本发明的各个方面可以实现为系统、方法或程序产品。因此，本发明的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、微代码等），或硬件和软件方面结合的实施方式，这里可

以统称为“电路”、“模块”或“系统”。

下面参照图 10 来描述根据本发明的这种实施方式的电子设备 1000。图 10 显示的电子设备 1000 仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

如图 10 所示，电子设备 1000 以通用计算设备的形式表现。电子设备 1000 的组件可以包括但不限于：上述至少一个处理单元 1010、上述至少一个存储单元 1020、连接不同系统组件（包括存储单元 1020 和处理单元 1010）的总线 1030。

其中，所述存储单元存储有程序代码，所述程序代码可以被所述处理单元 1010 执行，使得所述处理单元 1010 执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的步骤。例如，所述处理单元 1010 可以执行如图 1 中所示步骤 S100：获取用户输入的数据字典描述；步骤 S110：将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；步骤 S120：获取所述数据字典描述对应的关键词；步骤 S130：基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；步骤 S140：基于所述数据结构确定数据项；步骤 S150：基于所述数据项、所述数据结构和所述数据流生成数据字典。

存储单元 1020 可以包括易失性存储单元形式的可读介质，例如随机存取存储单元（RAM）10201 和/或高速缓存存储单元 10202，还可以进一步包括只读存储单元（ROM）10203。

存储单元 1020 还可以包括具有一组（至少一个）程序模块 10205 的程序/实用工具 10204，这样的程序模块 10205 包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

总线 1030 可以为表示几类总线结构中的一种或多种，包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

电子设备 1000 也可以与一个或多个外部设备 500（例如键盘、指向设备、蓝牙设备等）通信，还可与一个或者多个使得用户能与该电子设备 1000 交互的设备通信，和/或使得该电子设备 1000 能与一个或多个其它计算设备进行通信的任何设备（例如路由器、调制解调器等等）通信。这种通信可以通过输入/输出（I/O）接口 1050 进行。并且，电子设备 1000 还可以通过网络适配器 1060 与一个或者多个网络（例如局域网（LAN），广域网（WAN）和/或公共网络，例如因特网）通信。如图所示，网络适配器 1060 通过总线 1030 与电子设备 1000 的其它模块通信。应当明白，尽管图中未示出，可以结合电子设备 1000 使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储系统等。

通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本发明实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、终端装置、或者网络设备等）执行根据本发明实施方式的方法。

在本发明的示例性实施例中，还提供了一种计算机可读存储介质，其上存储有能够实现本说明书上述方法的程序产品。在一些可能的实施方式中，本发明的各个方面还可以实现为一种程序产品的形式，其包括程序代码，当所述程序产品在终端设备上运行时，所述

程序代码用于使所述终端设备执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的步骤。

参考图 11 所示,描述了根据本发明的实施方式的用于实现上述方法的程序产品 1100,其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码,并可以在终端设备,例如个人电脑上运行。然而,本发明的程序产品不限于此,在本文件中,非易失性可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

所述程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM 或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

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

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

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

此外,上述附图仅是根据本发明示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,上述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其他实施例。本申请旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本发明未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由权利要求指出。

## 权利要求

1、一种数据字典生成方法，其特征在于，所述方法包括：

获取用户输入的数据字典描述；

5 将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；

获取所述数据字典描述对应的关键词；

基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；

基于所述数据结构确定数据项；

10 基于所述数据项、所述数据结构和所述数据流生成数据字典。

2、根据权利要求1所述的方法，其特征在于，在所述将所述数据字典描述输入预设的意图识别模型之前还包括：

获取预先设置的数据字典描述样本集合；

识别所述数据字典描述样本集合中的数据字典描述样本对应的意图信息；

15 将所述数据字典描述样本输入所述意图识别模型，由所述意图识别模型输出意图信息，将所述意图识别模型输出的意图信息和识别出的所述数据字典描述样本对应的意图信息进行比对，如不一致，则调整所述意图识别模型的参数，直至所述意图识别模型输出的意图信息与识别出的所述数据字典描述样本对应的意图信息一致。

20 3、根据权利要求1所述的方法，其特征在于，所述获取所述数据字典描述对应的关键词包括：

将所述数据字典描述分句；

将所述数据字典描述分成的句子与预设的数据字典描述句模板库中的数据字典描述句模板进行比对，以确定与所述数据字典描述分成的句子相匹配的数据字典描述句模板；

25 按照所述数据字典描述句模板中规定的数据字典描述关键词的位置，确定所述分成的句子中的关键词。

4、根据权利要求1所述的方法，其特征在于，所述获取所述数据字典描述对应的关键词包括：

获取预先设置的数据字典描述样本集合；

确定所述数据字典描述样本集合中的数据字典描述样本的关键词；

30 将所述数据字典描述样本输入第一机器学习模型，由所述第一机器学习模型输出关键词，将所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词进行比对，如不一致，则调整所述第一机器学习模型的参数，直至所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词一致；

35 将所述数据字典描述输入所述第一机器学习模型，获取由所述第一机器学习模型输出所述数据字典描述对应的关键词。

5、根据权利要求4所述的方法，其特征在于，所述基于所述关键词和所述意图信息

生成所述数据字典描述对应的数据流和数据结构包括:

获取预先设置的数据字典描述对应的关键词和意图信息样本集合;

确定所述数据字典描述对应的关键词和意图信息样本对应的数据流和数据结构;

5 将所述数据字典描述对应的关键词和意图信息样本输入所述第二机器学习模型,由所述第二机器学习模型输出数据流与数据结构,将所述第二机器学习模型输出的数据流与数据结构与确定的数据流与数据结构进行比对,如不一致,则调整所述第二机器学习模型的参数,直至所述第二机器学习模型输出的数据流和数据结构与确定的数据流与数据结构一致;

10 将所述数据流和数据结构输入所述第二机器学习模型,获取由所述第二机器学习模型输出所述数据流和数据结构。

6、根据权利要求1所述的方法,其特征在于,所述基于所述数据结构确定数据项包括:

基于所述数据结构确定数据项名;

基于所述数据项名在预设数据库中匹配相应的数据项。

15 7、根据权利要求4所述的方法,其特征在于,在所述基于所述数据项、所述数据结构和所述数据流生成数据字典之后包括:

将所述数据字典存储至共享池,并赋予相应的等级;

获取用户登录信息;

20 基于所述用户登录信息中包含的该用户的权限信息和该用户所要调取的数据字典的等级,确定所述用户是否能调取相应的数据字典。

8、一种数据字典生成装置,其特征在于,包括:

第一获取模块,用于获取用户输入的数据字典描述;

25 第二获取模块,用于将所述数据字典描述输入预设的意图识别模型,获取由所述意图识别模型输出的所述数据字典描述对应的意图信息,所述意图信息表示需要生成的数据字典的用途;

第三获取模块,获取所述数据字典描述对应的关键词;

第一生成模块,根据所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构;

确定模块,用于基于所述数据结构确定数据项;

30 第二生成模块,用于基于所述数据项、所述数据结构和所述数据流生成数据字典。

9、根据权利要求8所述的装置,其特征在于,所述装置还包括:

35 意图模型训练模块,用于,获取预先设置的数据字典描述样本集合;识别所述数据字典描述样本集合中的数据字典描述样本对应的意图信息;将所述数据字典描述样本输入所述意图识别模型,由所述意图识别模型输出意图信息,将所述意图识别模型输出的意图信息和识别出的所述数据字典描述样本对应的意图信息进行比对,如不一致,则调整所述意图识别模型的参数,直至所述意图识别模型输出的意图信息与识别出的所述数据字典描述样本对应的意图信息一致。

10、根据权利要求 8 所述的装置，其特征在于，所述第三获取模块还可配置为：

将所述数据字典描述分句；将所述数据字典描述分成的句子与预设的数据字典描述句模板库中的数据字典描述句模板进行比对，以确定与所述数据字典描述分成的句子相匹配的数据字典描述句模板；按照所述数据字典描述句模板中规定的数据字典描述关键词的位置，确定所述分成的句子中的关键词。11、根据权利要求 8 所述的装置，其特征在于，所述第三获取模块还可配置为：

获取预先设置的数据字典描述样本集合；确定所述数据字典描述样本集合中的数据字典描述样本的关键词；将所述数据字典描述样本输入第一机器学习模型，由所述第一机器学习模型输出关键词，将所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词进行比对，如不一致，则调整所述第一机器学习模型的参数，直至所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词一致；将所述数据字典描述输入所述第一机器学习模型，获取由所述第一机器学习模型输出所述数据字典描述对应的关键词。

12、根据权利要求 8 所述的装置，其特征在于，所述第一生成模块装置还可配置为：

获取预先设置的数据字典描述对应的关键词和意图信息样本集合，确定所述数据字典描述对应的关键词和意图信息样本对应的数据流和数据结构，将所述数据字典描述对应的关键词和意图信息样本输入所述第二机器学习模型，由所述第二机器学习模型输出数据流与数据结构，将所述第二机器学习模型输出的数据流与数据结构与确定的数据流与数据结构进行比对，如不一致，则调整所述第二机器学习模型的参数，直至所述第二机器学习模型输出的数据流和数据结构与确定的数据流与数据结构一致，将所述数据流和数据结构输入所述第二机器学习模型，获取由所述第二机器学习模型输出所述数据流和数据结构。

13、根据权利要求 8 所述的装置，其特征在于，所述确定模块还可配置为：

基于所述数据结构确定数据项名；基于所述数据项名在预设数据库中匹配相应的数据项。

14、根据权利要求 8 所述的装置，其特征在于，所述装置还包括：

共享模块，用于：将所述数据字典存储至共享池，并赋予相应的等级；获取用户登录信息；基于所述用户登录信息中包含的该用户的权限信息和该用户所要调取的数据字典的等级，确定所述用户是否能调取相应的数据字典。

15、一种数据字典生成的电子设备，其特征在于，包括存储器，配置为存储可执行指令，处理器，配置为执行存储器中存储的可执行指令，所述计算机可读指令被所述处理器执行时，所述处理器用于执行以下处理：

获取用户输入的数据字典描述；

将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；

获取所述数据字典描述对应的关键词；

基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；

基于所述数据结构确定数据项；

基于所述数据项、所述数据结构和所述数据流生成数据字典。

16、根据权利要求 15 所述的电子设备，其特征在于，在所述将所述数据字典描述输

入预设的意图识别模型之前，所述处理器用于实现以下步骤：

获取预先设置的数据字典描述样本集合；

识别所述数据字典描述样本集合中的数据字典描述样本对应的意图信息；

5 将所述数据字典描述样本输入所述意图识别模型，由所述意图识别模型输出意图信息，  
将所述意图识别模型输出的意图信息和识别出的所述数据字典描述样本对应的意图信息  
进行比对，如不一致，则调整所述意图识别模型的参数，直至所述意图识别模型输出的意  
图信息与识别出的所述数据字典描述样本对应的意图信息一致。

17、根据权利要求 15 所述的电子设备，其特征在于，所述处理器通过执行如下处理  
来实现所述获取所述数据字典描述对应的关键词：

10 将所述数据字典描述分句；

将所述数据字典描述分成的句子与预设的数据字典描述句模板库中的数据字典描述  
句模板进行比对，以确定与所述数据字典描述分成的句子相匹配的数据字典描述句模板；

按照所述数据字典描述句模板中规定的数据字典描述关键词的位置，确定所述分成的  
句子中的关键词。

15 18、根据权利要求 15 所述的电子设备，其特征在于，所述处理器通过执行如下处理  
来实现所述获取所述数据字典描述对应的关键词：

获取预先设置的数据字典描述样本集合；

确定所述数据字典描述样本集合中的数据字典描述样本的关键词；

20 将所述数据字典描述样本输入第一机器学习模型，由所述第一机器学习模型输出关键  
词，将所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词进行  
比对，如不一致，则调整所述第一机器学习模型的参数，直至所述第一机器学习模型输出  
的关键词与确定的所述数据字典描述样本的关键词一致；

将所述数据字典描述输入所述第一机器学习模型，获取由所述第一机器学习模型输出  
所述数据字典描述对应的关键词。

25 19、根据权利要求 15 所述的电子设备，其特征在于，所述处理器通过执行如下处理  
来实现所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构：

获取预先设置的数据字典描述对应的关键词和意图信息样本集合；

确定所述数据字典描述对应的关键词和意图信息样本对应的数据流和数据结构；

30 将所述数据字典描述对应的关键词和意图信息样本输入所述第二机器学习模型，由所  
述第二机器学习模型输出数据流与数据结构，将所述第二机器学习模型输出的数据流与数  
据结构与确定的数据流与数据结构进行比对，如不一致，则调整所述第二机器学习模型的  
参数，直至所述第二机器学习模型输出的数据流和数据结构与确定的数据流与数据结构一  
致；

35 将所述数据流和数据结构输入所述第二机器学习模型，获取由所述第二机器学习模型  
输出所述数据流和数据结构。

20、根据权利要求 15 所述的电子设备，其特征在于，所述处理器通过执行如下处理  
来实现所述基于所述数据结构确定数据项：

基于所述数据结构确定数据项名；

基于所述数据项名在预设数据库中匹配相应的数据项。

21、根据权利要求 15 所述的电子设备，其特征在于，在所述基于所述数据项、所述数据结构和所述数据流生成数据字典之后，所述处理器于执行以下步骤：

将所述数据字典存储至共享池，并赋予相应的等级；

5 获取用户登录信息；

基于所述用户登录信息中包含的该用户的权限信息和该用户所要调取的数据字典的等级，确定所述用户是否能调取相应的数据字典。

22、一种计算机非易失性可读存储介质，其特征在于，所述计算机程序被处理器执行时，使得所述处理器用于实现以下步骤：获取用户输入的数据字典描述；

10 将所述数据字典描述输入预设的意图识别模型，获取由所述意图识别模型输出的所述数据字典描述对应的意图信息，所述意图信息表示需要生成的数据字典的用途；

获取所述数据字典描述对应的关键词；

基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构；

基于所述数据结构确定数据项；

15 基于所述数据项、所述数据结构和所述数据流生成数据字典。

23、根据权利要求 22 所述的计算机非易失性可读存储介质，其特征在于，在所述将所述数据字典描述输入预设的意图识别模型之前，所述处理器用于实现以下步骤：获取预先设置的数据字典描述样本集合；

识别所述数据字典描述样本集合中的数据字典描述样本对应的意图信息；

20 将所述数据字典描述样本输入所述意图识别模型，由所述意图识别模型输出意图信息，将所述意图识别模型输出的意图信息和识别出的所述数据字典描述样本对应的意图信息进行比对，如不一致，则调整所述意图识别模型的参数，直至所述意图识别模型输出的意图信息与识别出的所述数据字典描述样本对应的意图信息一致。

24、根据权利要求 22 所述的计算机非易失性可读存储介质，其特征在于，所述获取  
25 所述数据字典描述对应的关键词，所述处理器用于实现以下步骤：

将所述数据字典描述分句；

将所述数据字典描述分成的句子与预设的数据字典描述句模板库中的数据字典描述句模板进行比对，以确定与所述数据字典描述分成的句子相匹配的数据字典描述句模板；

30 按照所述数据字典描述句模板中规定的数据字典描述关键词的位置，确定所述分成的句子中的关键词。

25、根据权利要求 22 所述的计算机非易失性可读存储介质，其特征在于，所述获取  
所述数据字典描述对应的关键词，所述处理器用于实现以下步骤：

获取预先设置的数据字典描述样本集合；

确定所述数据字典描述样本集合中的数据字典描述样本的关键词；

35 将所述数据字典描述样本输入第一机器学习模型，由所述第一机器学习模型输出关键词，将所述第一机器学习模型输出的关键词与确定的所述数据字典描述样本的关键词进行比对，如不一致，则调整所述第一机器学习模型的参数，直至所述第一机器学习模型输出

的关键词与确定的所述数据字典描述样本的关键词一致；

将所述数据字典描述输入所述第一机器学习模型，获取由所述第一机器学习模型输出所述数据字典描述对应的关键词。

- 5 26、根据权利要求 22 所述的计算机非易失性可读存储介质，其特征在于，所述基于所述关键词和所述意图信息生成所述数据字典描述对应的数据流和数据结构，所述处理器用于实现以下步骤：

获取预先设置的数据字典描述对应的关键词和意图信息样本集合；

确定所述数据字典描述对应的关键词和意图信息样本对应的数据流和数据结构；

- 10 将所述数据字典描述对应的关键词和意图信息样本输入所述第二机器学习模型，由所述第二机器学习模型输出数据流与数据结构，将所述第二机器学习模型输出的数据流与数据结构与确定的数据流与数据结构进行比对，如不一致，则调整所述第二机器学习模型的参数，直至所述第二机器学习模型输出的数据流和数据结构与确定的数据流与数据结构一致；

- 15 将所述数据流和数据结构输入所述第二机器学习模型，获取由所述第二机器学习模型输出所述数据流和数据结构。

27、根据权利要求 22 所述的计算机非易失性可读存储介质，其特征在于，所述基于所述数据结构确定数据项，所述处理器用于实现以下步骤：

基于所述数据结构确定数据项名；

基于所述数据项名在预设数据库中匹配相应的数据项。

- 20 28、根据权利要求 22 所述的计算机非易失性可读存储介质，其特征在于，在所述基于所述数据项、所述数据结构和所述数据流生成数据字典之后，所述处理其用于执行以下步骤：

将所述数据字典存储至共享池，并赋予相应的等级；

获取用户登录信息；

- 25 基于所述用户登录信息中包含的该用户的权限信息和该用户所要调取的数据字典的等级，确定所述用户是否能调取相应的数据字典。

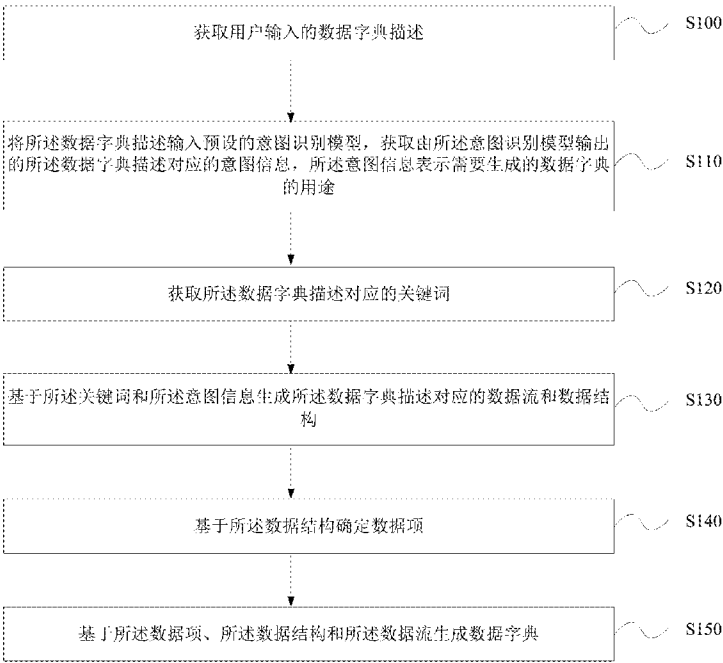


图 1

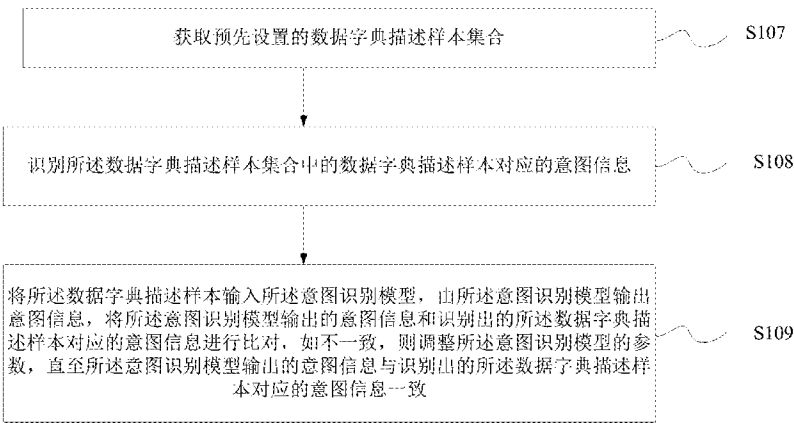


图 2

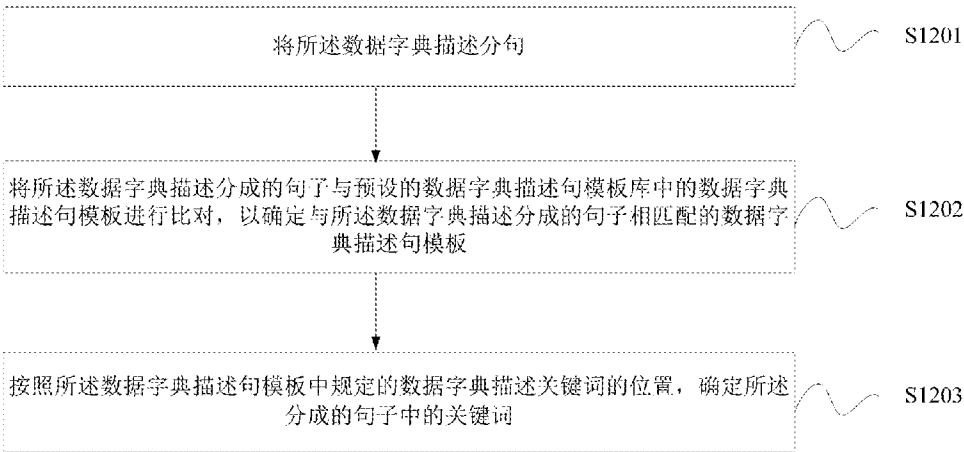


图 3

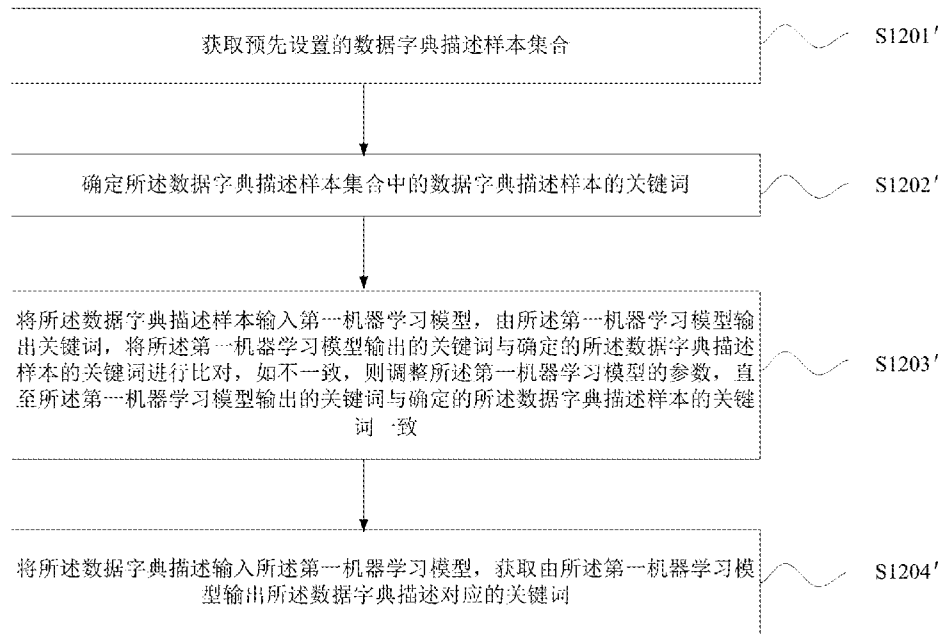


图 4

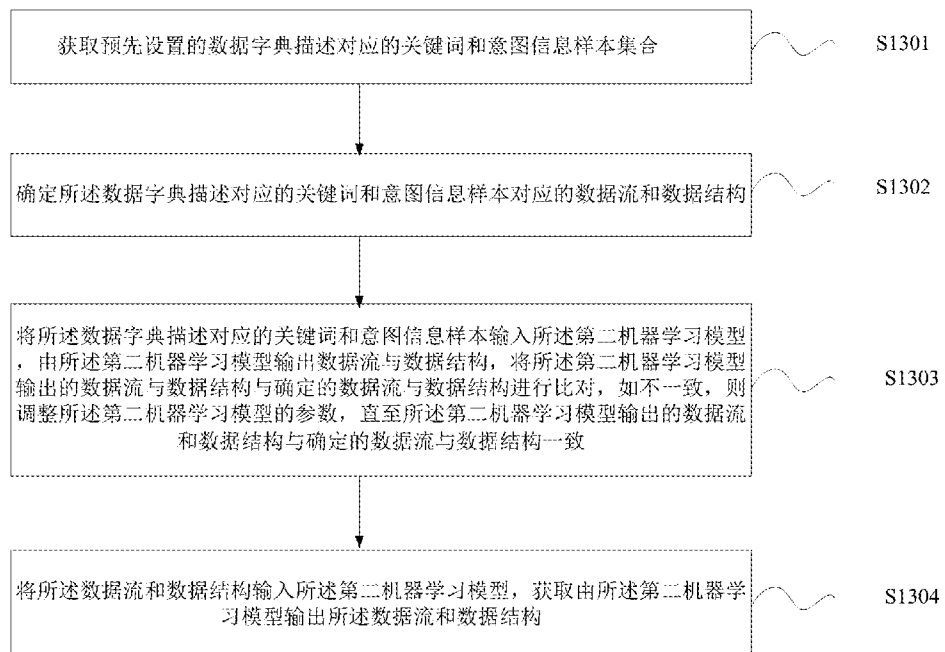


图 5

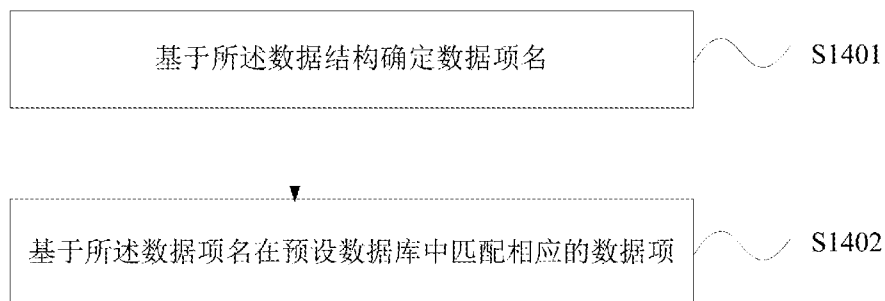


图 6

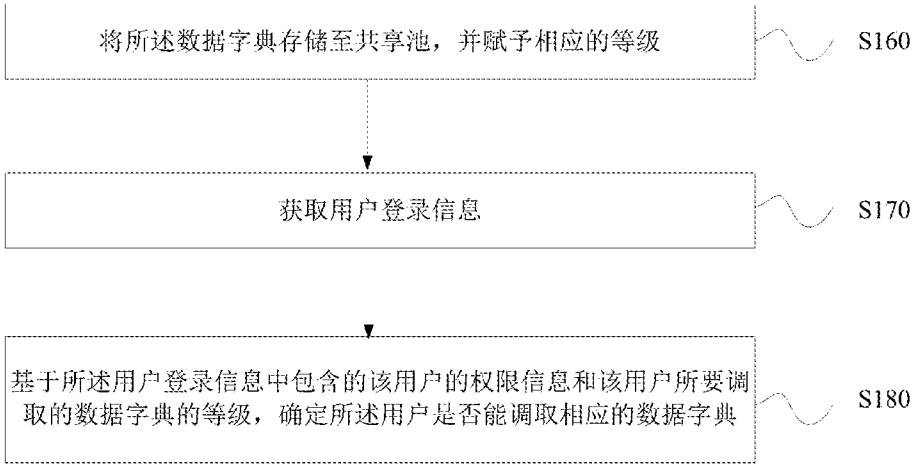


图 7

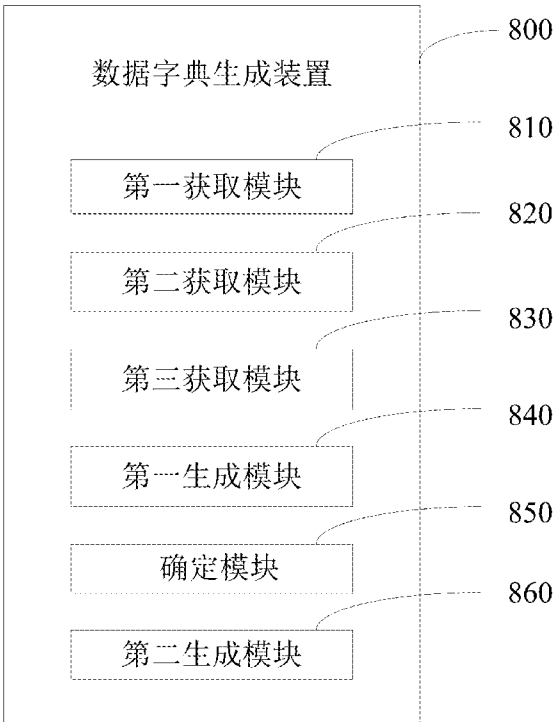


图 8

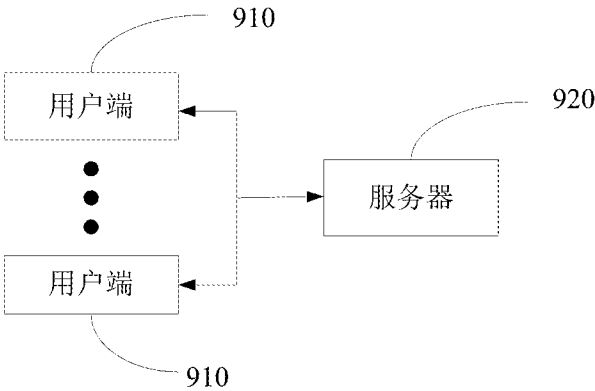


图 9

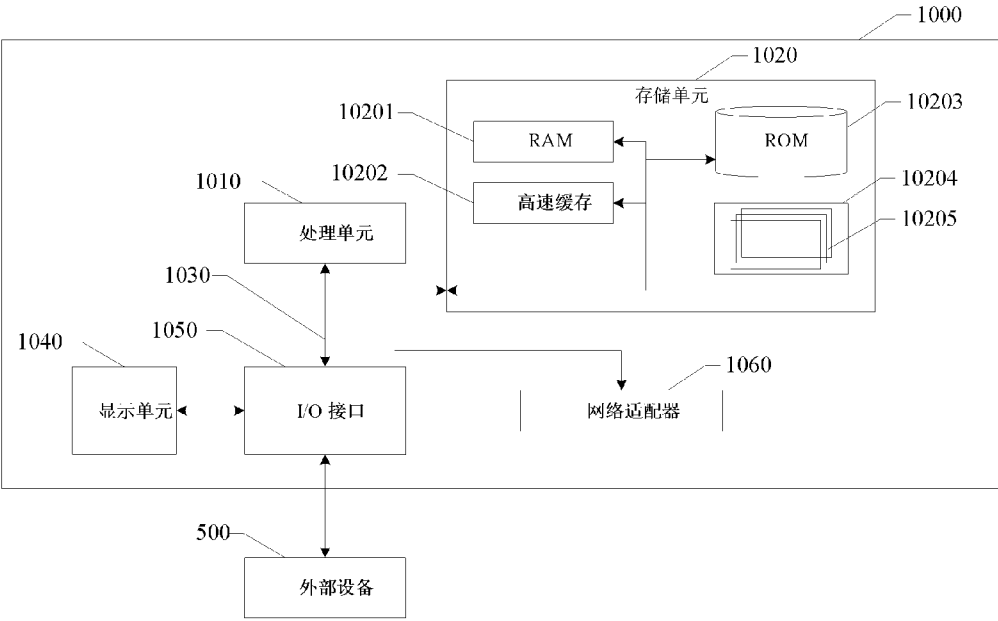


图 10

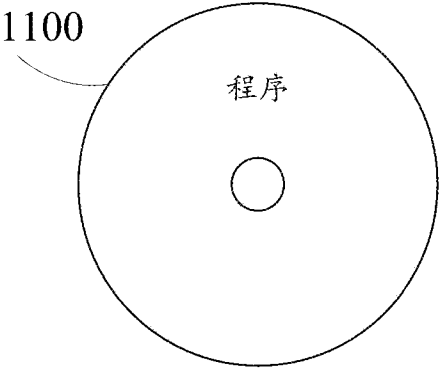


图 11

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/103434

**A. CLASSIFICATION OF SUBJECT MATTER**

G06N 20/00(2019.01)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)

G06N, 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)

CNKI, CNPAT, WPI, EPODOC: 数据, 字典, 辞典, 词典, 生成, 建立, 构建, 产生, 关系表, 系统表, 关系, 结构, 用途, 意图, 目的, 分析, 解析, 描述, 关键词, 关键字, 模型, data, dictionary, generat+, creat+, build+, mak+, produc+, conceiv+, relation, table, system, purpos+, aim, intent+, analy+, parse+, key, word?, model

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                    | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| A         | CN 101673287 A (KINGDEE SOFTWARE (CHINA) CO., LTD.) 17 March 2010 (2010-03-17)<br>description, page 4, paragraph 1 to page 10, the last paragraph, page 15, paragraph 5 (from the bottom) to page 16, paragraph 3, and figures 1-2, 4 | 1-28                  |
| A         | CN 105005592 A (YONYOU UP INFORMATION TECHNOLOGY CO., LTD.) 28 October 2015 (2015-10-28)<br>entire document                                                                                                                           | 1-28                  |
| A         | CN 103927353 A (BEIJING NETQIN TECHNOLOGY CO., LTD.) 16 July 2014 (2014-07-16)<br>entire document                                                                                                                                     | 1-28                  |
| A         | CN 108280099 A (GUANGZHOU DONG JING COMPUTER SCIENCE & TECHNOLOGY CO., LTD.) 13 July 2018 (2018-07-13)<br>entire document                                                                                                             | 1-28                  |
| A         | CN 104850566 A (JURONG ZHONGXIN SOFTWARE SCIENCE & TECHNOLOGY CO., LTD.) 19 August 2015 (2015-08-19)<br>entire document                                                                                                               | 1-28                  |



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

19 January 2020

Date of mailing of the international search report

10 February 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

International application No.

PCT/CN2019/103434

### C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages            | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------|-----------------------|
| A         | US 2015205787 A1 (LOGAWI DATA ANALYTICS, L.L.C.) 23 July 2015 (2015-07-23)<br>entire document | 1-28                  |

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

International application No.

**PCT/CN2019/103434**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|--------------------------------------|
| CN                                        | 101673287  | A  | 17 March 2010                        | None                    |                                      |
| CN                                        | 105005592  | A  | 28 October 2015                      | None                    |                                      |
| CN                                        | 103927353  | A  | 16 July 2014                         | None                    |                                      |
| CN                                        | 108280099  | A  | 13 July 2018                         | None                    |                                      |
| CN                                        | 104850566  | A  | 19 August 2015                       | None                    |                                      |
| US                                        | 2015205787 | A1 | 23 July 2015                         | None                    |                                      |

## 国际检索报告

国际申请号

PCT/CN2019/103434

## A. 主题的分类

G06N 20/00 (2019.01) 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)

G06N, 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))

CNKI, CNPAT, WPI, EPDOC: 数据, 字典, 辞典, 词典, 生成, 建立, 构建, 产生, 关系表, 系统表, 关系, 结构, 用途, 意图, 目的, 分析, 解析, 描述, 关键词, 关键字, 模型, data, dictionary, generat+, creat+, build+, mak+, produc+, conceiv+, relation, table, system, purpos+, aim, intent+, analy+, parse+, key, word?, model

## C. 相关文件

| 类 型* | 引用文件, 必要时, 指明相关段落                                                                                         | 相关的权利要求 |
|------|-----------------------------------------------------------------------------------------------------------|---------|
| A    | CN 101673287 A (金蝶软件中国有限公司) 2010年 3月 17日 (2010 - 03 - 17)<br>说明书第4页第1段至10页最后1段, 第15页倒数第5段至第16页第3段、图1-2, 4 | 1-28    |
| A    | CN 105005592 A (用友优普信息技术有限公司) 2015年 10月 28日 (2015 - 10 - 28)<br>全文                                        | 1-28    |
| A    | CN 103927353 A (北京网秦天下科技有限公司) 2014年 7月 16日 (2014 - 07 - 16)<br>全文                                         | 1-28    |
| A    | CN 108280099 A (广州市动景计算机科技有限公司) 2018年 7月 13日 (2018 - 07 - 13)<br>全文                                       | 1-28    |
| A    | CN 104850566 A (句容中新软件科技有限公司) 2015年 8月 19日 (2015 - 08 - 19)<br>全文                                         | 1-28    |
| A    | US 2015205787 A1 (LOGAWI DATA ANALYTICS, LLC) 2015年 7月 23日 (2015 - 07 - 23)<br>全文                         | 1-28    |

☐ 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

“&amp;” 同族专利的文件

国际检索实际完成的日期

2020年 1月 19日

国际检索报告邮寄日期

2020年 2月 10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

戚颖

电话号码 86-10-53961718

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2019/103434

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|----------------|
| CN          | 101673287  | A  | 2010年 3月 17日   | 无    |                |
| CN          | 105005592  | A  | 2015年 10月 28日  | 无    |                |
| CN          | 103927353  | A  | 2014年 7月 16日   | 无    |                |
| CN          | 108280099  | A  | 2018年 7月 13日   | 无    |                |
| CN          | 104850566  | A  | 2015年 8月 19日   | 无    |                |
| US          | 2015205787 | A1 | 2015年 7月 23日   | 无    |                |