

(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 4 月 23 日 (23.04.2020)



(10) 国际公布号
WO 2020/077896 A1

(51) 国际专利分类号:
G06F 16/332 (2019.01) **G06F 16/36** (2019.01)

(21) 国际申请号: PCT/CN2019/070844

(22) 国际申请日: 2019 年 1 月 8 日 (08.01.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811204606.1 2018 年 10 月 16 日 (16.10.2018) CN

(71) 申请人: 深圳壹账通智能科技有限公司
(**ONE CONNECT SMART TECHNOLOGY CO.,LTD.**
(**SHENZHEN**)) [CN/CN]; 中国广东省深圳市
前海深港合作区前湾一路 1 号 A 栋 201
室, Guangdong 518052 (CN)。

(72) 发明人: 臧磊 (**ZANG, Lei**); 中国广东省深圳市
前海深港合作区前湾一路 1 号 A 栋 201 室,
Guangdong 518052 (CN)。 傅婧 (**FU, Jing**); 中国
广东省深圳市前海深港合作区前湾一路 1 号 A 栋
201 室, Guangdong 518052 (CN)。 郭鹏程 (**GUO,**
Pengcheng); 中国广东省深圳市前海深港合作区前
湾一路 1 号 A 栋 201 室, Guangdong 518052 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (**ADVANCE CHINA IP LAW OFFICE**); 中国
广东省广州市天河区珠江东路 6 号 4501 房
(部位: 自编 01-03 和 08-12 单元) (仅限办公
用途), Guangdong 510623 (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,

(54) **Title:** METHOD AND APPARATUS FOR GENERATING QUESTION DATA, COMPUTER DEVICE, AND STORAGE
MEDIUM

(54) 发明名称: 提问数据生成方法、装置、计算机设备和存储介质

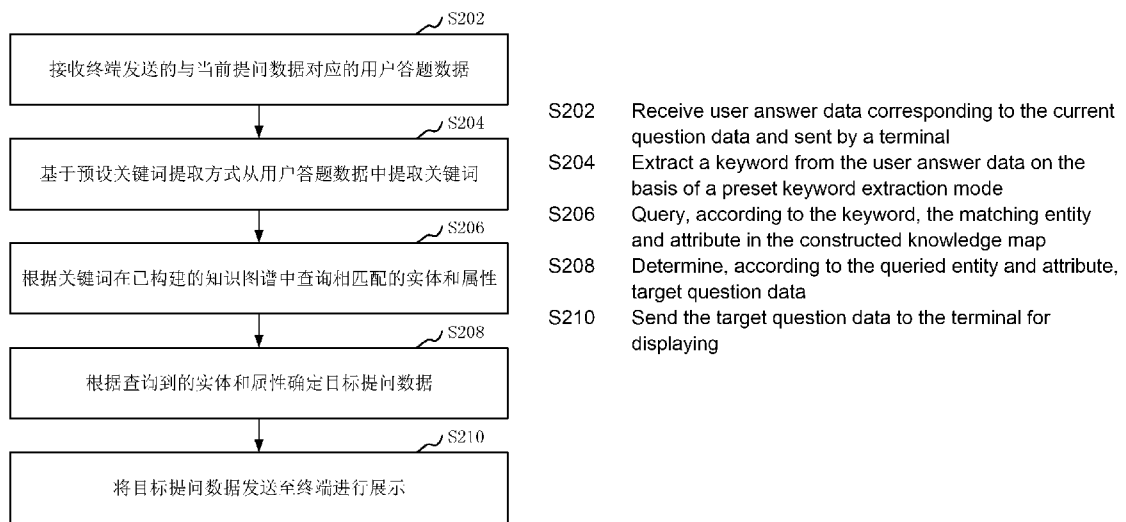


图 2

(57) **Abstract:** A question data generation method, comprising: receiving user answer data corresponding to the current question data and sent by a terminal; extracting a keyword from the user answer data on the basis of a preset keyword extraction mode; querying, according to the keyword, the matching entity and attribute in the constructed knowledge map; determining, according to the queried entity and attribute, target question data; and sending the target question data to the terminal for displaying.

(57) **摘要:** 一种提问数据生成方法, 包括: 接收终端发送的与当前提问数据对应的用户答题数据; 基于预设关键词提取方式从所述用户答题数据中提取关键词; 根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性; 根据查询到的实体和属性确定目标提问数据; 将所述目标提问数据发送至所述终端进行展示。

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

提问数据生成方法、装置、计算机设备和存储介质

本申请要求于 2018 年 10 月 16 日提交中国专利局，申请号为 2018112046061，申请名称为“提问数据生成方法、装置、计算机设备和存储介质”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。

5

技术领域

本申请涉及一种提问数据生成方法、装置、计算机设备和存储介质。

背景技术

10 传统信贷场景下的问答以人工为主，随着计算机技术的不断发展，逐渐出现了系统性的人机问答，由终端代替业务员向用户提问，降低了针对业务员的行业培训成本。然而，系统性人机问答通常基于预先配置的固定题库模板对应确定提问数据，即信贷面审过程中，服务器根据固定题库模板对应生成一套固定的提问数据，并通过终端将该固定的提问数据依次进行展示。

15 然而，发明人意识到，目前，信贷面审过程中的提问数据是固定的，不具备动态性，可能导致信贷面审题库的泄露，从而存在信贷面审题库的安全性低的问题。

发明内容

20 根据本申请公开的各种实施例，提供一种提问数据生成方法、装置、计算机设备和存储介质。

一种提问数据生成方法包括：

接收终端发送的与当前提问数据对应的用户答题数据；

基于预设关键词提取方式从所述用户答题数据中提取关键词；

根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性；

25 根据查询到的实体和属性确定目标提问数据；及

将所述目标提问数据发送至所述终端进行展示。

一种提问数据生成装置包括：

接收模块，用于接收终端发送的与当前提问数据对应的用户答题数据；

提取模块，用于基于预设关键词提取方式从所述用户答题数据中提取关键词；

30 查询模块，用于根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性；

生成模块，用于根据查询到的实体和属性确定目标提问数据；及

发送模块，用于将所述目标提问数据发送至所述终端进行展示。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中储存有计算机可读

指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器实现本申请任意一个实施例中提供的提问数据生成方法的步骤。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器实现本申请任意一个实施例中提供的提问数据生成方法的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，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它的附图。

图 1 为根据一个或多个实施例中提问数据生成方法的应用场景图。

图 2 为根据一个或多个实施例中提问数据生成方法的流程示意图。

图 3 为另一个实施例中提问数据生成方法的流程示意图。

图 4 为根据一个或多个实施例中提问数据生成装置的框图。

图 5 为另一个实施例中提问数据生成装置的框图。

图 6 为根据一个或多个实施例中计算机设备的框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请提供的提问数据生成方法，可以应用于如图 1 所示的应用环境中。终端 102 通过网络与服务器 104 通过网络进行通信。服务器 104 接收终端 102 针对当前提问数据对应反馈的用户答题数据，从该用户答题数据中提取关键词，根据所提取的关键词在以构建的知识图谱中查询与该关键词相匹配的实体和属性，根据所查询到的实体和属性生成目标提问数据，并将该目标提问数据发送至终端 102 进行展示。终端 102 可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

在其中一个实施例中，如图 2 所示，提供了一种提问数据生成方法，以该方法应用于图 1 中的服务器为例进行说明，包括以下步骤：

S202，接收终端发送的与当前提问数据对应的用户答题数据。

当前提问数据是当前向用户提问的题目数据，即当前问答数据处理过程所对应的提问数据。当前提问数据具体可以是指离当前时间最近的提问数据，也即最新的提问数据。提

问数据具体可以通过显示屏进行展示的文本数据，也可以通过语音播报的方式展示的语音信息。当前提问数据与终端当前反馈的用户答题数据和用户答题影像想对应。用户答题数据是用户针对当前提问数据对应反馈的答题数据。用户答题数据具体可以是用户在终端手动录入的文本形式的答题数据，也可以是用户回答当前提问数据时终端对应采集到的语音形式的答题信息。

具体地，终端接收到服务器发送的当前提问数据时，通过语音播报或显示屏显示的方式向用户展示所接收到的当前提问数据，并实时检测用户针对该当前提问数据对应反馈用户答题数据。当检测到用户针对当前提问数据对应反馈的用户答题数据时，终端将所检测到的用户答题数据发送至服务器。

在其中一个实施例中，服务器接收问答指令，根据所接收到的问答指令获取相应的当前提问数据，将所获取到的当前提问数据发送至终端。服务器根据问答指令从预存储的预设题库中选择当前提问数据。服务器也可根据问答指令从预存储的提问素材集中获取相应的提问素材，根据所获取到的提问素材生成相应的当前提问数据。服务器还可根据问答指令从已构建的知识图谱中选择实体和相应的属性，并根据所选择的实体和属性对应生成当前数据。预设题库是由预先生成的多个提问数据组成的提问数据集。提问素材集是由多个提问素材组成的素材集合。提问素材比如身份证号、籍贯、现居地等。

在其中一个实施例中，终端向用户展示当前提问数据时，实时检测用户针对当前提问数据的答题操作，根据所检测到的答题操作获取相应的用户答题数据。答题操作比如用户在终端的录入操作，或用户的口述操作。

在其中一个实施例中，当接收到的用户答题数据为语音形式的答题数据时，服务器对该用户答题数据进行语音识别，获得与该用户答题数据对应的答题语音文本，根据所提取到的语音答题文本执行下述提问数据生成方法的相关步骤。语音识别可基于现有的语音识别技术，在此不再赘述。

S204，基于预设关键词提取方式从用户答题数据中提取关键词。

预设关键词提取方式是预先设定的用于提取用户答题数据中的关键词的方式。预设关键词提取方式可以是基于预设关键词库进行关键词匹配的方式；也可以是对用户答题数据进行分词，并按照预设筛选条件从分词获得的多个词中筛选关键词；还可以是基于已训练的关键词提取模型来进行关键词提取，关键词提取模型具体可分为基于有监督的、半监督的和无监督的训练方式训练获得的模型。

具体地，服务器基于预设关键词提取方式对所接收到的用户答题数据进行关键词提取，得到该用户答题数据对应的关键词。服务器对用户答题数据进行分词，将分词得到的各个词分别与预设关键词库进行关键词匹配，并将与预设关键词库匹配成功的词作为关键词。预设关键词库是由预先设定的多个关键词组成的关键词词集。

在其中一个实施例中，服务器对用户答题数据进行分词，得到分词结果，然后对分词结果中的各个词计算特征权重，并根据特征权重对分词结果中的各个词进行排序，进而根

据排序结果选取关键词。具体地，服务器对分词结果中的各个词计算 TF-IDF 权重，根据 TF-IDF 权重对各个词进行降序排列，并选取排序靠前的预设数量的词作为关键词。各个词的 TF-IDF 权重计算过程包括：首先，计算分词结果中的各个词的词频，可参考如下公式进行计算：

- 5 词频 TF=某个词在文档中出现的次数/文档的总词数；
 然后，计算各个词的逆文档词频 IDF，可参考如下公式进行计算：

$$\text{逆文档词频 IDF} = \log \left(\frac{\text{语料库的文档总数}}{\text{包含该词的文档数}+1} \right);$$

最后，计算词频 TF 与逆文档词频 IDF 的乘积得到 TF-IDF 权重。

- 10 在其中一个实施例中，服务器将用户答题数据输入预先训练好的关键词提取模型进行预测，获取与该用户答题数据对应的关键词。关键词提取模型是根据预先获取的训练样本集进行模型训练获得的模型。

S206，根据关键词在已构建的知识图谱中查询相匹配的实体和属性。

- 15 知识图谱是一张包含各种实体或概念及其关系的语义网络图。知识图谱中的节点为与实体对应的实体节点，与任一实体节点连接的边表示该任一实体节点所对应的实体的属性，该与任一实体节点连接的边的另一端所连接的实体节点所对应的实体为该任一实体节点对应的属性值。比如，该任一实体节点对应的实体为 XX 集团总部，与该任一实体节点连接的边表示地址，则该与 XX 集团总部连接的边的另一端所连接的实体节点所对应的实体为深圳市福田区。知识图谱中的实体可包括概念、人名、地名和企业机构名称等。知识图谱中的属性可为该属性对应的边所连接的两个实体之间的关系（实体关系），或者相应
20 实体本身所具备的特性。

可以理解的是，知识图谱具体可以是包含大量三元组的知识库。三元组包括实体、实体关系，以及通过实体关系与该实体相关联的关联实体。将各个三元组中的实体关系作为实体与相应关联实体之间的属性，根据属性在实体和对应的关联实体之间建立相应的有向连接，从而构建相应的知识图谱。

- 25 具体地，服务器将提取到的关键词分别与已构建的知识图谱中的各个实体和属性进行匹配，将与关键词匹配成功的实体和属性确定为与该关键词相匹配的实体和属性。从用户答题数据中提取到的关键词有多个，服务器将所提取到的各个关键词分别与知识图谱中的各个实体进行匹配，将匹配成功的实体确定为根据关键词对应查询到的实体。类似地，服务器将所提取到的各个关键词分别与知识图谱中的各个属性进行匹配，将匹配成功的属性
30 确定为根据关键词对应查询到的属性。

在其中一个实施例中，服务器可将所提取到的关键词的同义词分别与知识图谱中的各个属性进行匹配，将匹配成功的属性确定为与该关键词相匹配的属性。服务器也可将所查询到的关键词分别与知识图谱中各个属性的关键词进行匹配，并将匹配成功的属性确定为与该关键词相匹配的属性。

在其中一个实施例中，服务器将所提取到的关键词与知识图谱中的各实体进行匹配，以将该关键词对应到知识图谱的某一个实体上，从而在知识图谱中找到与该关键词相匹配的实体。服务器通过 NLP(Natural Language Processing，自然语言处理)技术理解用户答题数据所表达的语义，并将所理解的语义与知识图谱中的属性进行匹配，以识别出与用户答题数据对应的属性，并将该属性确定为与相应关键词相匹配的属性。服务器也可将当前提问数据所表达的语义与知识图谱中的属性进行匹配，以确定与所提取到的关键词相匹配的属性。

在其中一个实施例中，从用户答题数据中提取到的关键词可理解为从该用户答题数据中识别出的命名实体，该用户答题数据所表达的语义可理解为该命名实体所对应的属性。命名实体包括人名、地名和企业机构名称等。当提取到的关键词有多个时，服务器可根据该多个关键词对应确定命名实体。

在其中一个实施例中，服务器根据所识别出的命名实体在知识图谱中查询相匹配的实体。服务器分别计算命名实体和知识图谱中各实体之间的匹配率，将匹配率达到预设阈值时的实体确定为与命名实体相匹配的实体。匹配率是指命名实体和相应实体之间的匹配程度。比如，命名实体为企业机构名称时，与该命名实体相匹配的实体可以是该企业机构名称，也可以是该企业机构名称对应的企业机构所对应的其他企业机构名称，如简称或全称等。命名实体为地名时，与该命名实体相匹配的实体可以是该地名，也可以是离该地名所对应的地址在预设范围内的其他地址所对应的其他地名。

举例说明，假设当前提问数据为“请问，你现居住地是哪里？”，对应获取到的用户答题数据为“我目前住在南山区科苑路 58 号”。则按照上述方式从该用户答题数据中识别出的命名实体为“南山区科苑路 58 号”，基于上述关键词提取方式或者语义理解方式可对应确定该命名实体所对应的属性为“地址”。服务器从知识图谱中查询到的与命名实体相匹配的实体，可以是“南山区科苑路 58 号”，或者“南山区科苑路 59 号”等。

在其中一个实施例中，在当前提问数据为是非问题时，服务器可将用户答题数据的内容文本确定为该用户答题数据所对应的关键词。进一步地，服务器可根据该当前提问数据和从用户答题数据中对应提取到的关键词，对应确定该用户答题数据所表示的语义，并根据该语义从已构建的知识图谱中查询与该语义相匹配的实体和属性。举例说明，假设当前提问数据为“南山区科苑路 59 号有个 XX 银行，是吗？”，对应的用户答题数据为“是的”，服务器根据该用户答题数据所表示的语义确定的实体和属性可分别为“南山区科苑路 57 号”和“地址”。

S208，根据查询到的实体和属性确定目标提问数据。

目标提问数据是指继当前提问数据之后，再一次向用户发起提问时的题目数据。目标提问数据具体可以是可通过显示屏进行展示的文本数据，也可以是可通过语音播报的方式展示的语音信息。

具体地，服务器根据所查询到的实体和属性，在相应知识图谱中查询通过该属性与该

实体相关联的关联实体。服务器根据所查询到的实体、属性和相应的关联实体生成相应的目标提问数据。

举例说明，基于上述例子，假设根据识别出的命名实体在知识图谱中对应查询到的实体为“南山区科苑路 59 号”，根据该命名实体所对应的属性“地址”，在知识图谱中对应查询到的关联实体为“XX 银行”，则对应生成的目标提问数据可以是“南山区科苑路 59 号有个 XX 银行，是吗？”，或者“南山区科苑路 58 号附近有个 XX 银行，是吗？”

在其中一个实施例中，服务器根据从用户答题数据中提取到的关键词，在已构建的知识图谱中查询与该关键词相匹配的实体，并根据该知识图谱对应确定与该实体相连接的一条或多条边。该一条或多条边表示该实体的属性，服务器可从该一个或多个属性中选择一个属性，并根据所选择的属性和相应实体生成目标提问数据。在其中一个实施例中，服务器在知识图谱中对应查询通过该一条或多条边与该实体关联的一个或多个关联实体。该一个或多个关联实体为属性所对应的属性值。服务器从该一个或多个属性中选择一个属性，以及从相应的一个或多个关联实体中选择与所选择的属性相对应的关联实体，并根据所选择的属性、关联实体和相应的实体对应生成目标提问数据。

举例说明，假设当前提问数据为“请问，你在哪个公司上班？”，对应获取到的用户答题数据为“我在 XX 集团上班”。则按照上述方式根据该用户答题数据中的关键词对应查询到的实体为“XX 集团”，基于该实体在知识图谱中对应查询到的属性可以是“地址”、“法人”、“成立时间”等。按照上述方式对应生成的目标提问数据可以是“XX 集团的法人是谁？”，或者“XX 集团的法人是张三，是吗？”等。

S210，将目标提问数据发送至终端进行展示。

具体地，服务器将根据所查询到的实体和属性对应确定的目标提问数据通过网络发送至终端。终端将所接收到的目标提问数据通过显示屏显示或语音播报的方式展示给相应用户。

在其中一个实施例中，终端通过语音播报或显示屏显示的方式向用户展示所接收到的目标提问数据时，实时检测用户针对该目标提问数据对应反馈用户答题数据，将所检测到的用户答题数据发送至服务器，以使得服务器根据所接收到的用户答题数据继续执行上述提问数据生成方法的相关步骤。

在其中一个实施例中，服务器将目标提问数据发送至用户终端，以通过该用户终端将目标提问数据展示给相应用户。服务器还可将目标提问数据发送至业务员终端进行展示。

上述提问数据生成方法，接收终端针对当前提问数据对应反馈的用户答题数据，基于预设关键词提取方式从该用户答题数据中提取相应的关键词，并根据所提取的关键词在已构建的知识图谱中对应查询与关键词相匹配的实体和属性，进而根据所查询到的实体和属性生成相应的目标提问数据，提高了目标提问数据的精准性。根据用户答题数据和已构建的知识图谱动态确定下一次提问时的目标提问数据，以保证目标提问数据的不确定性，

能够有效避免由各提问数据构成的题库泄露的问题，从而提高了题库的安全性。

在其中一个实施例中，步骤 S208 包括：根据查询到的实体和属性，在知识图谱中查询基于属性与实体相关联的关联实体；关联实体有多个；从各关联实体中选取目标关联实体；根据实体、属性和目标关联实体生成目标提问数据。

5 关联实体是在知识图谱中通过边（属性）与另一实体相关联的实体。可以理解的是，实体和关联实体均为知识图谱中的实体节点，实体和关联实体通过属性相关联，也即属性用于表现实体和相应关联实体之间的实体关系。通过属性相关联的两个实体可互为关联实体。关联实体可以是相应实体所对应的属性值，也可以是通过属性所表征的实体关系与相应实体相关联的另一实体。

10 具体地，服务器根据所查询到的实体和属性，在知识图谱中查询到多个通过该属性与该实体相关联的关联实体。服务器从所查询到的多个关联实体中选取目标关联实体，并根据该目标关联实体，以及相应的实体和属性对应生成目标提问数据。

在其中一个实施例中，服务器可从所查询到的多个关联实体中随机选择一个关联实体作为目标关联实体。服务器也可获取与所接收到的用户答题数据对应的用户标识，统计所
15 查询到的关联实体的总数量，并对该用户标识按照统计的总数量计算哈希值，进而根据计算所得的哈希值从该多个关联实体中选择目标关联实体。服务器还可通过轮询的方式依次从所查询到的多个关联实体中选择一个关联实体作为目标关联实体。

上述实施例中，从所查询到的多个关联实体中选择目标关联实体，进而根据目标关联实体以及相应的实体和属性对应生成目标提问数据，进一步提高了目标提问数据的不确定性，从而进一步提高了题库的安全性。
20

在其中一个实施例中，步骤 S204 包括：对用户答题数据进行预处理；按照基于预设关键词库的匹配方式，从预处理后的用户答题数据中提取关键词。

预处理包括分词处理和去停用词处理。分词处理是将文本形式的用户答题数据划分成一个一个的词的过程。分词处理对应的分词算法有多种，比如，基于字符串匹配的分词算法、基于语义分析的分词算法或者基于统计的分词算法等。基于字符串匹配的分词算法如
25 正向最大匹配算法、逆向最大匹配算法、最少切分算法或者双向最大匹配算法。已训练好的分词模型也可用于对用户答题数据进行分词处理。分词模型具体可以是隐马尔可夫模型或 CRF（conditional random field algorithm，条件随机场算法）模型等。停用词是指在信息检索中，为节省存储空间和提高检索效率，在处理自然语言数据（或文本）之前或之后会自动过滤掉的某些字或词，比如语气助词、客套词、介词或连接词等，比如的、吗、呢、啊等。预设关键词库是由预先设置的多个预设关键词构成的关键词词集。预设关键词库中
30 还可包括各个预设关键词的同义词。

具体地，服务器对所接收到的用户答题数据进行分词处理，得到分词结果，并对分词结果进行去停用词处理，获得相应的候选关键词集。服务器可将分词结果中的各个词分别
35 与预先构建的停用词库进行匹配，将与停用词库匹配成功的词确定为停用词，并从分词结

果中去除该停用词。服务器将去除停用词后的分词结果中的各个词作为候选关键词，获得相应的候选关键词集。进一步地，服务器将该候选关键词集中的各个候选关键词，分别与预设关键词库中的预设关键词，以及预设关键词的同义词进行匹配。当匹配成功时，表明匹配成功的候选关键词为预设关键词库中的预设关键词或该预设关键词的同义词，服务器

5 将该候选关键词确定为从预处理后的用户答题数据中提取的关键词。

在其中一个实施例中，服务器采用 NLP(Natural Language Processing 自然语言处理) 技术对用户答题数据进行处理，并对处理后的用户答题数据进行语义分析与理解，获得相应的关键词。在其中一个实施例中，服务器对用户答题数据进行预处理，并按照基于词典和正则的匹配方式从预处理后的用户答题数据中提取关键词。具体地，服务器将预处理获得

10 的各个候选关键词分别与预设词典中的预设关键词进行匹配，并将匹配成功的候选关键词确定为所提取到的关键词。

上述实施例中，基于预设关键词库提取关键词，使得从用户答题数据中所提取到的关键词更加准确。

在其中一个实施例中，步骤 S206 之前，上述提问数据生成方法还包括：获取目标数据；识别目标数据中的各目标实体和目标实体之间的实体关系；根据各目标实体和相应的实体关系，按照预设构建方式构建知识图谱。

15

目标数据是用于构建知识图谱的原始数据。目标数据具体可以是一个或多个指定领域的原始数据，指定领域比如金融领域和制造业领域等。以金融领域的原始数据为例，金融领域可包括银行、证券、保险、信托、信贷和基金等行业，每个行业可包括多个企业机构，每个企业机构对应应有相应的企业原始数据，企业原始数据可包括内部数据和外部数据等，内部数据是存储在企业本地数据库的数据，比如运营数据和业务数据等，外部数据是可通过第三方平台从其他存储空间爬取的数据，比如工商数据、银监数据、人民银行数据和公司年报等。集团企业的企业原始数据还可包括各个子公司的原始数据，比如投资关系、成立时间

20 和地址等。对于包含银行的企业原始数据，还可包括各个支行的地址等。

具体地，服务器接收到知识图谱的构建指令时，根据所接收到的构建指令从各种渠道获取已被公开的原始数据，并从本地数据库中获取预存储的且未被公开的原始数据，获得相应的目标数据。服务器对所获取到的目标数据进行数据加工，进而识别出该目标数据中的各个目标实体和各个目标实体之间的实体关系。服务器将任意两个目标实体之间的实体关系作为该任意两个目标实体之间的属性，并根据该属性在该任意两个实体之间建立相应的有向连接。对于从目标数据中识别出的各个目标实体和相应的实体关系，服务器均执行

30 上述步骤，从而构建与该目标数据对应的知识图谱。

在其中一个实施例中，服务器可使用数据仓库技术对所获取到的目标数据进行数据加工。服务器将按照上述方式构建的知识图谱中的关系型数据保存到图数据库中。该图数据库可以是 Neo4j (一个高性能的 NOSQL 图形数据库) 数据库。

35 在其中一个实施例中，目标数据包括多个指定领域的原始数据。对于每个指定领域的

原始数据，服务器可分别以领域作为知识图谱的顶点（核心节点），根据从目标数据中识别出的实体关系对应确定与该核心节点相连接的关关节点。类似地，服务器根据实体关系分别确定各个关关节点所连接的关关节点，并将各个节点之间的实体关系作为属性，根据所确定的各个节点和相应的属性构建知识图谱。核心节点和关关节点均为知识图谱中的节点（实体节点）。在其中一个实施例中，服务器可按照上述方式分别构建各个指定领域所对应的知识图谱，并根据任意两个指定领域之间存在交集的实体或属性，对该任意两个指定领域各自对应的知识图谱进行融合，获得跨领域的知识图谱。任意两个指定领域之间存在交集的实体或属性是指该任意两个指定领域各自对应的知识图谱中均包含的实体或属性。

上述实施例中，根据目标数据识别相应的实体和实体关系，进而根据识别出的实体和实体关系构建相应的知识图谱，以便于在动态生成提问数据时可以直接使用该已构建的知识图谱，提高提问数据的生成效率。

在其中一个实施例中，用户答题数据包括答题语音信息；步骤 S204 包括：对答题语音信息进行语音识别获得相应的答题语音文本；基于预设关键词提取方式从答题语音文本中提取关键词。

答题语音信息是用户针对当前提问数据进行答题操作时对应采集到的语音信息。答题语音文本是答题语音信息中所携带的语音文本内容。

具体地，服务器从所接收到的用户答题数据中获取相应的答题语音信息，并对所获取到的答题语音信息进行语音识别，以从该答题语音信息中提取相应的答题语音文本。服务器对提取出的答题语音文本进行分词，并将分词得到的各个词分别与预设关键词库进行关键词匹配，将与预设关键词库匹配成功的词作为从答题语音文本中提取到的关键词。

在其中一个实施例中，服务器通过预先训练好的语音识别模型从答题语音信息中提取相应的答题语音文本，并通过预先训练好的关键词提取模型从该答题语音文本中提取相应的关键词。可以理解的是，服务器还可基于其他现有技术中的语音识别技术从答题语音信息中提取相应的答题语音文本，类似地，还可基于其他现有技术中的关键词提取技术对应提取关键词，在此不再赘述。

在其中一个实施例中，服务器对答题语音信息进行声纹识别，以提取相应的目标声纹特征，并根据该目标声纹特征对针对当前提问数据进行答题操作的用户进行身份认证。当身份认证通过时，服务器继续执行对答题语音信息进行语音识别获得相应的答题语音文本的步骤。

上述实施例中，当用户答题数据中包括答题语音信息时，基于语音识别技术识别出相应的答题语音文本，进而从该答题语音文本中对应提取关键词，提高了用户答题数据的处理效率，从而提高了提问数据的生成效率。

在其中一个实施例中，用户答题数据包括用户答题文本和用户答题影像；步骤 S206 之前，上述提问数据生成方法还包括：确定与用户答题文本对应的答题分值，并确定与用

户答题影像对应的表情分值；根据答题分值和表情分值确定相应的综合分值；当综合分值低于预设分值阈值时，执行根据关键词在已构建的知识图谱中查询相匹配的实体和属性的步骤。

5 用户答题文本是文本形式的答题数据。终端可根据用户的手动录入操作获取相应的用户答题文本，还可将从用户进行答题操作时对应采集到的语音信息中识别出的语音内容作为用户答题文本。用户答题影像是在用户针对当前提问数据对应反馈答题数据时对应采集到的影像信息。影像信息具体可以是由影像采集器采集到的以图像或视频等形式存在的信息。用户答题影像可包括但不限于用户答题图像和用户答题视频。用户答题图像比如用户答题时对应采集到的包括用户人脸图像的图像。用户答题视频比如用户答题过程中对应
10 采集到的视频，即用户进行答题操作过程中采集到的视频。影像采集器可以是摄像头，摄像头可以是配置于终端的摄像头，也可以是与终端点对点连接的独立部件。用户答题影像与用户答题文本相对应。

具体地，服务器从所接收到的用户答题数据中获取相应的用户答题文本和用户答题影像。服务器按照预设答题分值确定方式确定与用户答题文本对应的答题分值，并按照预设
15 表情分值确定方式确定与用户答题影像对应的表情分值。服务器根据所确定的答题分值和表情分值确定与相应用户答题数据对应的综合分值，并将所确定的综合分值与预设分值阈值进行比较。当综合分值低于预设分值阈值时，服务器执行根据从该用户答题数据所包含的用户答题文本中提前到的关键词，在已构建的知识图谱中查询相匹配的实体和属性的步骤。

20 预设答题分值确定方式可以是根据用户答题文本与相应预设答案之间的匹配率，以及当前提问数据所对应的预设答题分值对应确定答题分值；也可以是将用户答题文本中的关键词与预设关键词进行匹配，以确定相应的答题分值；还可以是将用户答题文本输入已训练的答题分值预测模型进行预测，获得相应的答题分值。预设表情分值确定方式可以是
25 从用户答题影像中提取相应的用户微表情，基于微表情识别技术根据提取到的用户微表情确定与用户答题影像对应的表情分值。

在其中一个实施例中，服务器根据用户答题影像获取多帧用户人脸图像，基于微表情识别技术分别确定该多帧用户人脸图像各自对应的表情分值，并根据该多个表情分值确定
30 与该用户答题影像对应的表情分值。在其中一个实施例中，当用户答题影像对应的表情分值低于预设表情分值阈值时，服务器执行根据相应的关键词在已构建的知识图谱中查询相匹配的实体和属性的步骤。

上述实施例中，当综合分值低于预设分值阈值时，表明用户答题文本的正确性存疑，则基于已构建的知识图谱选择与当前提问数据相关的实体和属性，并动态生成相应的目标提问数据，以使得动态生成的目标提问数据更准确。

在其中一个实施例中，用户答题数据与用户标识对应；上述提问数据生成方法还包括：
35 统计与用户标识对应的提问数据总数量；当统计的总数量达到预设总数量时，停止提问数

据的生成流程；根据与用户标识对应的各提问数据所对应的综合分值，确定与用户标识对应的总分值；将总分值推送至终端进行展示。

具体地，服务器接收终端针对用户标识和当前提问数据对应反馈的用户答题数据，统计与该用户标识对应的提问数据总数量，并将统计的总数量与预设总数量进行比较。当统计的总数量达到预设总数量时，服务器则停止当前的提问数据生成流程。服务器按照上述方式确定当前提问数据所对应的综合分值，并获取与相应用户标识对应的、已有的各个提问数据所对应的综合分值，对当前确定的综合分值和对应获取到的多个综合分值进行直接求和或者加权求和，以确定与相应用户标识对应总分值，并将该总分值发送至终端进行展示。

在其中一个实施例中，服务器根据计算所得的总分值和预设总分值阈值生成相应的提示信息，并将该提示信息推送至终端。比如上述方法应用于信贷面审过程中的问答环节时，服务器可对应生成表示面审成功或失败的提示信息。面审是指在信贷业务办理过程中对用户身份进行审核。上述方法应用于面试环节时，服务器可对应生成表示面试成功或失败的提示信息。

在其中一个实施例中，服务器按照上述方式确定当前提问数据对应的综合分值，并根据该综合分值更新相应用户标识所对应的已有总分值。当更新后的总分值达到预设总分值阈值时，服务器停止当前的提问数据生成流程。

如图3所示，在其中一个实施例中，提供了一种提问数据生成方法，该方法具体包括以下步骤：

S302，接收终端发送的与当前提问数据对应的用户答题数据。

S304，对用户答题数据进行预处理。

S306，按照基于预设关键词库的匹配方式，从预处理后的用户答题数据中提取关键词。

S308，获取目标数据。

S310，识别目标数据中的各目标实体和目标实体之间的实体关系。

S312，根据各目标实体和相应的实体关系，按照预设构建方式构建知识图谱。

S314，根据关键词在已构建的知识图谱中查询相匹配的实体和属性。

S316，根据查询到的实体和属性，在知识图谱中查询基于属性与实体相关联的关联实体；关联实体有多个。

S318，从各关联实体中选取目标关联实体。

S320，根据实体、属性和目标关联实体生成目标提问数据。

S322，将目标提问数据发送至终端进行展示。

上述实施例中，从与当前提问数据对应的用户答题数据中提取关键词，并基于已构建的知识图谱根据关键词查询相应匹配的实体和属性，对应确定关联实体，进而根据实体、属性和相应的关联实体动态生成目标提问数据，并发送至终端进行展示，提高了目标提问数据的不确定性，从而提高了题库的安全性。

在其中一个实施例中，上述提问数据生成方法可应用于信贷面审的问答环节，也可以应用于面试环节，还可以应用于公安或海关的稽查审问的问答环节等。

应该理解的是，虽然图 2-3 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-3 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 4 所示，提供了一种提问数据生成装置 400，包括：接收模块 402、提取模块 404、查询模块 406、生成模块 408 和发送模块 410，其中：

接收模块 402，用于接收终端发送的与当前提问数据对应的用户答题数据。

提取模块 404，用于基于预设关键词提取方式从用户答题数据中提取关键词。

查询模块 406，用于根据关键词在已构建的知识图谱中查询相匹配的实体和属性。

生成模块 408，用于根据查询到的实体和属性确定目标提问数据。

发送模块 410，用于将目标提问数据发送至终端进行展示。

在其中一个实施例中，生成模块 408，还用于根据查询到的实体和属性，在知识图谱中查询基于属性与实体相关联的关联实体；关联实体有多个；从各关联实体中选取目标关联实体；根据实体、属性和目标关联实体生成目标提问数据。

在其中一个实施例中，提取模块 404，还用于对用户答题数据进行预处理；按照基于预设关键词库的匹配方式，从预处理后的用户答题数据中提取关键词。

如图 5 所示，在其中一个实施例中，上述提问数据生成装置 400，还包括：知识图谱构建模块 412；

知识图谱构建模块 412，用于获取目标数据；识别目标数据中的各目标实体和目标实体之间的实体关系；根据各目标实体和相应的实体关系，按照预设构建方式构建知识图谱。

在其中一个实施例中，用户答题数据包括答题语音信息；提取模块 404，还用于对答题语音信息进行语音识别获得相应的答题语音文本；基于预设关键词提取方式从答题语音文本中提取关键词。

在其中一个实施例中，用户答题数据包括用户答题文本和用户答题影像；上述提问数据生成装置 400，还包括：确定模块 414；

确定模块 414，用于确定与用户答题文本对应的答题分值，并确定与用户答题影像对应的表情分值；根据答题分值和表情分值确定相应的综合分值；当综合分值低于预设分值阈值时，使得查询模块 406 执行根据关键词在已构建的知识图谱中查询相匹配的实体和属性的步骤。

在其中一个实施例中，用户答题数据与用户标识对应；确定模块 414，还用于统计与用户标识对应的提问数据总数量；当统计的总数量达到预设总数量时，停止提问数据的生成流程；根据与用户标识对应的各提问数据所对应的综合分值，确定与用户标识对应的总分值；将总分值推送至终端进行展示。

5 关于提问数据生成装置的具体限定可以参见上文中对于提问数据生成方法的限定，在此不再赘述。上述提问数据生成装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

10 在其中一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 6 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为易失性存储介质中的操作系统和计算机可读指令的运行提供环
15 境。该计算机设备的数据库用于存储已构建的知识图谱。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种提问数据生成方法。

本领域技术人员可以理解，图 6 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。
20

一种计算机设备，包括存储器和一个或多个处理器，存储器中储存有计算机可读指令，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器实现本申请任意一个实施例中提供的提问数据生成方法的步骤。

25 一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器实现本申请任意一个实施例中提供的提问数据生成方法的步骤。

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
30 过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）
35 或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为

说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

- 5 以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种提问数据生成方法，包括：

接收终端发送的与当前提问数据对应的用户答题数据；

基于预设关键词提取方式从所述用户答题数据中提取关键词；

根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性；

5 根据查询到的实体和属性确定目标提问数据；及

将所述目标提问数据发送至所述终端进行展示。

2、根据权利要求 1 所述的方法，其特征在于，所述根据查询到的实体和属性确定目标提问数据，包括：

10 根据查询到的实体和属性，在所述知识图谱中查询基于所述属性与所述实体相关联的关联实体；所述关联实体有多个；

从各所述关联实体中选取目标关联实体；及

根据所述实体、所述属性和所述目标关联实体生成目标提问数据。

3、根据权利要求 1 所述的方法，其特征在于，所述基于预设关键词提取方式从所述用户答题数据中提取关键词，包括：

15 对所述用户答题数据进行预处理；及

按照基于预设关键词库的匹配方式，从预处理后的用户答题数据中提取关键词。

4、根据权利要求 1 所述的方法，其特征在于，所述根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性之前，所述方法还包括：

获取目标数据；

20 识别所述目标数据中的各目标实体和所述目标实体之间的实体关系；及

根据各所述目标实体和相应的所述实体关系，按照预设构建方式构建知识图谱。

5、根据权利要求 1 至 4 任意一项所述的方法，其特征在于，所述用户答题数据包括答题语音信息；所述基于预设关键词提取方式从所述用户答题数据中提取关键词，包括：

对所述答题语音信息进行语音识别获得相应的答题语音文本；及

25 基于预设关键词提取方式从所述答题语音文本中提取关键词。

6、根据权利要求 1 至 4 任意一项所述的方法，其特征在于，所述用户答题数据包括用户答题文本和用户答题影像；所述根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性之前，所述方法还包括：

30 确定与所述用户答题文本对应的答题分值，并确定与所述用户答题影像对应的表情分值；

根据所述答题分值和所述表情分值确定相应的综合分值；及

当所述综合分值低于预设分值阈值时，执行所述根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性的步骤。

7、根据权利要求 6 所述的方法，其特征在于，所述用户答题数据与用户标识对应；

所述方法还包括：

统计与所述用户标识对应的提问数据总数量；

当统计的所述总数量达到预设总数量时，停止提问数据的生成流程；

5 根据与所述用户标识对应的各提问数据所对应的综合分值，确定与所述用户标识对应的总分值；及
将所述总分值推送至所述终端进行展示。

8、一种提问数据生成装置，包括：

接收模块，用于接收终端发送的与当前提问数据对应的用户答题数据；

提取模块，用于基于预设关键词提取方式从所述用户答题数据中提取关键词；

10 查询模块，用于根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性；

生成模块，用于根据查询到的实体和属性确定目标提问数据；及

发送模块，用于将所述目标提问数据发送至所述终端进行展示。

9、一种计算机设备，包括存储器和一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

接收终端发送的与当前提问数据对应的用户答题数据；

基于预设关键词提取方式从所述用户答题数据中提取关键词；

根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性；

根据查询到的实体和属性确定目标提问数据；及

20 将所述目标提问数据发送至所述终端进行展示。

10、根据权利要求 9 所述的计算机设备，其特征在于，所述根据查询到的实体和属性确定目标提问数据，包括：

根据查询到的实体和属性，在所述知识图谱中查询基于所述属性与所述实体相关联的关联实体；所述关联实体有多个；

25 从各所述关联实体中选取目标关联实体；及

根据所述实体、所述属性和所述目标关联实体生成目标提问数据。

11、根据权利要求 9 所述的计算机设备，其特征在于，所述基于预设关键词提取方式从所述用户答题数据中提取关键词，包括：

对所述用户答题数据进行预处理；及

30 按照基于预设关键词库的匹配方式，从预处理后的用户答题数据中提取关键词。

12、根据权利要求 9 所述的计算机设备，其特征在于，所述计算机可读指令被所述处理器执行时，使得所述处理器在执行所述根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性之前，还执行以下步骤：

获取目标数据；

35 识别所述目标数据中的各目标实体和所述目标实体之间的实体关系；及

根据各所述目标实体和相应的所述实体关系，按照预设构建方式构建知识图谱。

13、根据权利要求 9 至 12 任意一项所述的计算机设备，其特征在于，所述用户答题数据包括答题语音信息；所述基于预设关键词提取方式从所述用户答题数据中提取关键词，包括：

- 5 对所述答题语音信息进行语音识别获得相应的答题语音文本；及
 基于预设关键词提取方式从所述答题语音文本中提取关键词。

14、根据权利要求 9 至 12 任意一项所述的计算机设备，其特征在于，所述用户答题数据包括用户答题文本和用户答题影像；所述计算机可读指令被所述处理器执行时，使得所述处理器在执行所述根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性
10 之前，还执行以下步骤：

 确定与所述用户答题文本对应的答题分值，并确定与所述用户答题影像对应的表情分值；

 根据所述答题分值和所述表情分值确定相应的综合分值；及

 当所述综合分值低于预设分值阈值时，执行所述根据所述关键词在已构建的知识图谱
15 中查询相匹配的实体和属性的步骤。

15、根据权利要求 14 所述的计算机设备，其特征在于，所述用户答题数据与用户标识对应；所述处理器执行所述计算机可读指令时还执行以下步骤：

 统计与所述用户标识对应的提问数据总数量；

 当统计的所述总数量达到预设总数量时，停止提问数据的生成流程；

20 根据与所述用户标识对应的各提问数据所对应的综合分值，确定与所述用户标识对应的总分值；及

 将所述总分值推送至所述终端进行展示。

16、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

25 接收终端发送的与当前提问数据对应的用户答题数据；

 基于预设关键词提取方式从所述用户答题数据中提取关键词；

 根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性；

 根据查询到的实体和属性确定目标提问数据；及

 将所述目标提问数据发送至所述终端进行展示。

30 17、根据权利要求 16 所述的存储介质，其特征在于，所述根据查询到的实体和属性确定目标提问数据，包括：

 根据查询到的实体和属性，在所述知识图谱中查询基于所述属性与所述实体相关联的关联实体；所述关联实体有多个；

 从各所述关联实体中选取目标关联实体；及

35 根据所述实体、所述属性和所述目标关联实体生成目标提问数据。

18、根据权利要求 16 所述的存储介质，其特征在于，所述基于预设关键词提取方式从所述用户答题数据中提取关键词，包括：

对所述用户答题数据进行预处理；及

按照基于预设关键词库的匹配方式，从预处理后的用户答题数据中提取关键词。

5 19、根据权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时，使得所述处理器在执行所述根据所述关键词在已构建的知识图谱中查询相匹配的实体和属性之前，还执行以下步骤：

获取目标数据；

识别所述目标数据中的各目标实体和所述目标实体之间的实体关系；及

10 根据各所述目标实体和相应的所述实体关系，按照预设构建方式构建知识图谱。

20、根据权利要求 16 至 19 任意一项所述的存储介质，其特征在于，所述用户答题数据包括答题语音信息；所述基于预设关键词提取方式从所述用户答题数据中提取关键词，包括：

对所述答题语音信息进行语音识别获得相应的答题语音文本；及

15 基于预设关键词提取方式从所述答题语音文本中提取关键词。

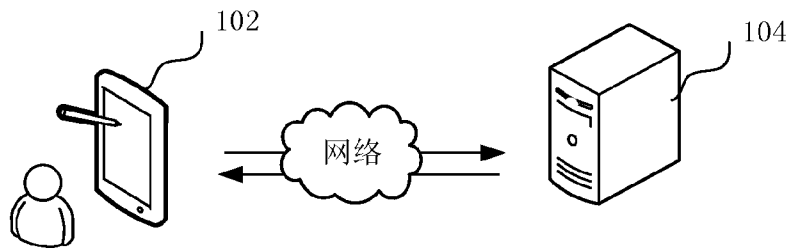


图 1

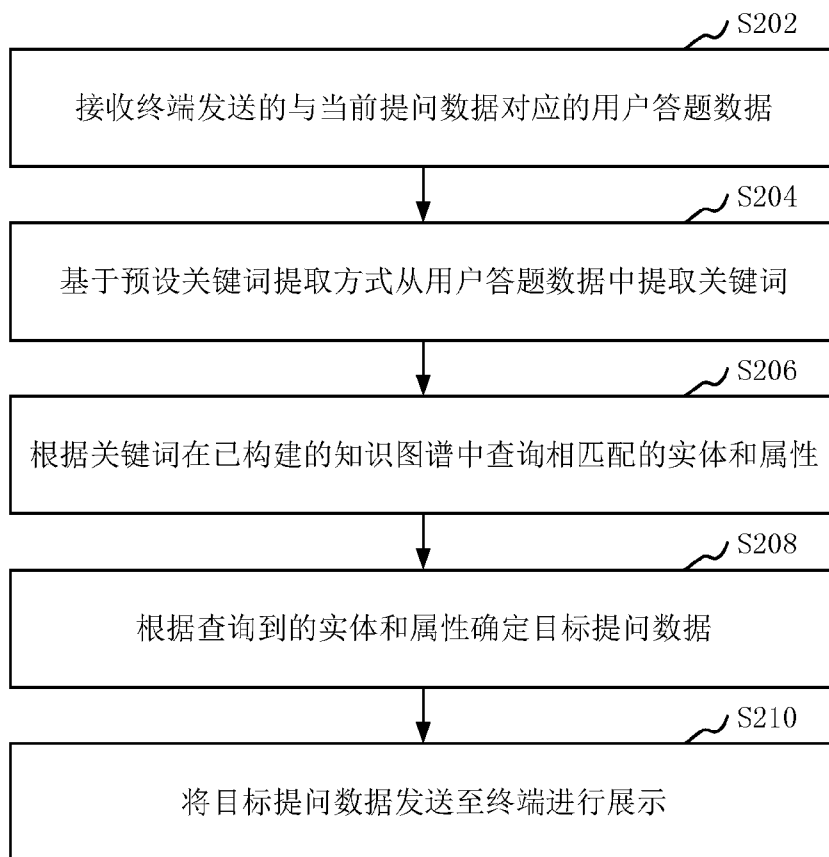


图 2

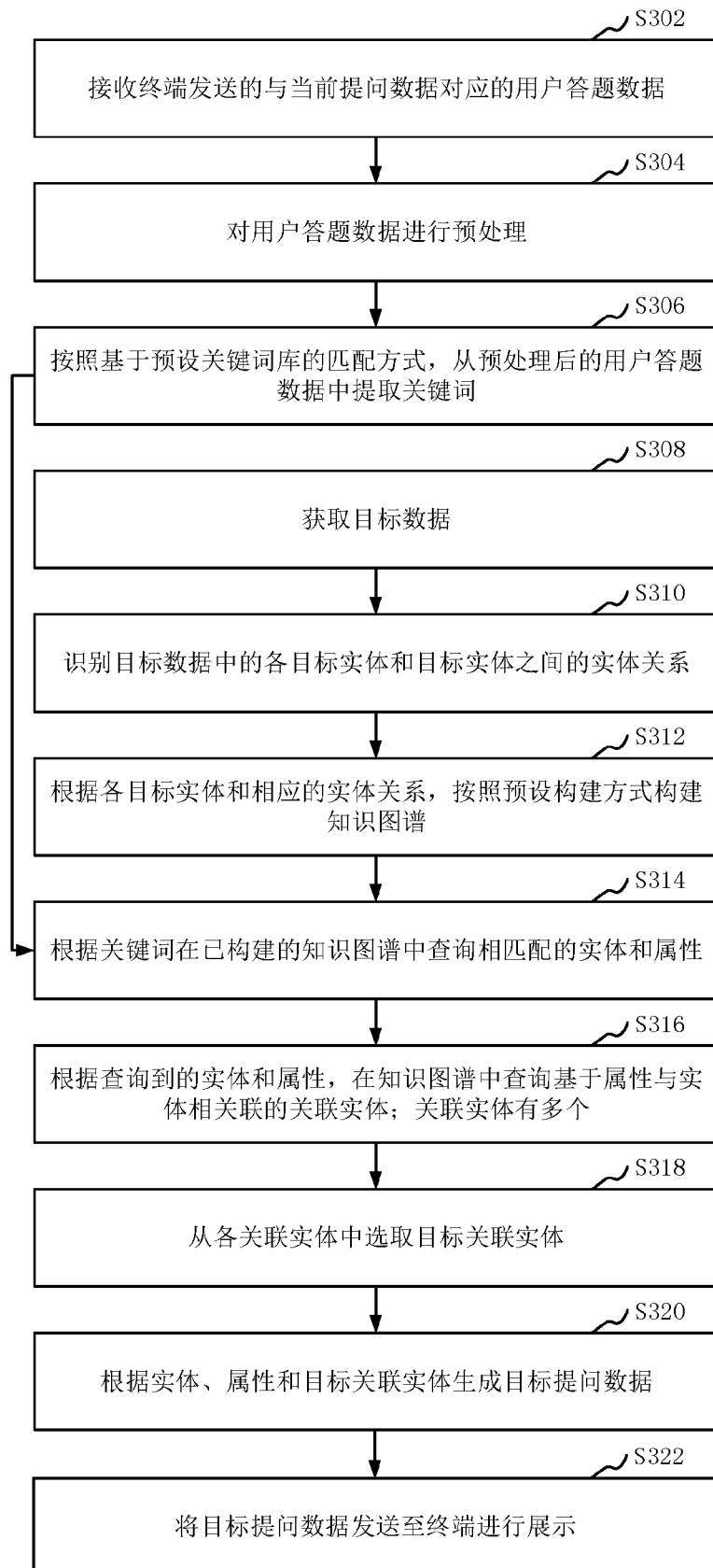


图 3

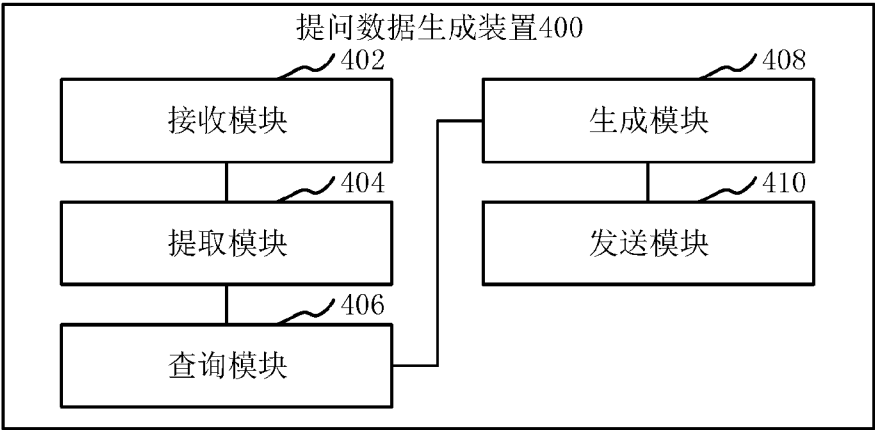


图 4

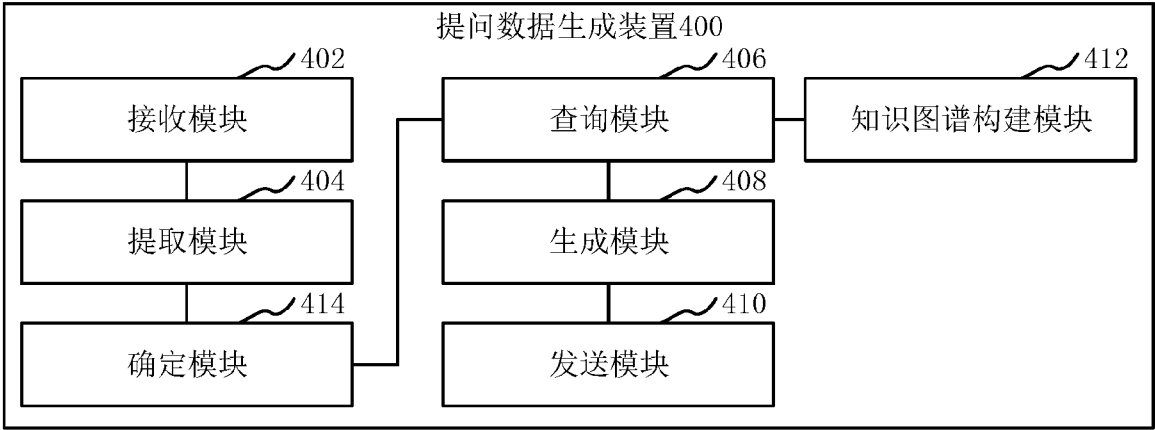


图 5

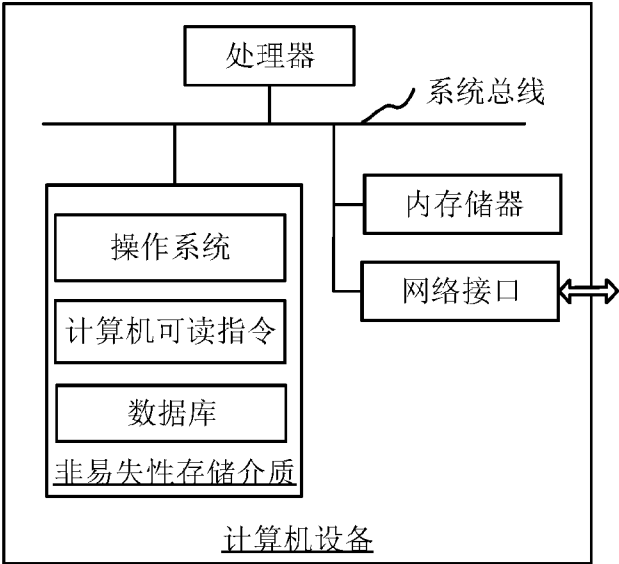


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/070844

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/332(2019.01)i; G06F 16/36(2019.01)n

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)

G06F16/-

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)

CNABS, CNTXT, CNKI, SIPOABS, DWPI: 回答, 答案, 回复, 答复, 对话, 生成, 产生, 确定, 提问, 问题; answer, reply, dialog, generate, produce, question

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108519998 A (BEIJING UNISOUND INFORMATION TECHNOLOGY CO., LTD.) 11 September 2018 (2018-09-11) claims 1-10, and description, paragraphs [0003]-[0115]	1-20
A	CN 101261690 A (BEIJING BAIWEN BAIDA NETWORK TECHNOLOGIES CO., LTD.) 10 September 2008 (2008-09-10) entire document	1-20
A	CN 104978396 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 14 October 2015 (2015-10-14) entire document	1-20
A	US 2014280087 A1 (IBM CORPORATION) 18 September 2014 (2014-09-18) 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

12 July 2019

Date of mailing of the international search report

25 July 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
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/CN2019/070844

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108519998	A	11 September 2018	None			
CN	101261690	A	10 September 2008	None			
CN	104978396	A	14 October 2015	None			
US	2014280087	A1	18 September 2014	US	9063975	B2	23 June 2015

国际检索报告

国际申请号

PCT/CN2019/070844

A. 主题的分类

G06F 16/332(2019.01)i; G06F 16/36(2019.01)n

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)

G06F16/-

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))

CNABS, CNTXT, CNKI, SIPOABS, DWPI:回答, 答案, 回复, 答复, 对话, 生成, 产生, 确定, 提问, 问题; answer, reply, dialog, generate, produce, question

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
X	CN 108519998 A (北京云知声信息技术有限公司) 2018年 9月 11日 (2018 - 09 - 11) 权利要求1-10, 说明书第[0003]-[0115]段	1-20
A	CN 101261690 A (北京百问百答网络技术有限公司) 2008年 9月 10日 (2008 - 09 - 10) 全文	1-20
A	CN 104978396 A (百度在线网络技术北京有限公司) 2015年 10月 14日 (2015 - 10 - 14) 全文	1-20
A	US 2014280087 A1 (IBM公司) 2014年 9月 18日 (2014 - 09 - 18) 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 7月 12日

国际检索报告邮寄日期

2019年 7月 25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

受权官员

甘文珍

电话号码 86-010-62411708

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/070844

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	108519998	A	2018年 9月 11日	无	
CN	101261690	A	2008年 9月 10日	无	
CN	104978396	A	2015年 10月 14日	无	
US	2014280087	A1	2014年 9月 18日	US 9063975 B2	2015年 6月 23日