(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
2021 年 8 月 19 日 (19.08.2021)

WIPO | PCT

(10) 国际公布号
WO 2021/159632 A1

(51) 国际专利分类号:
G06F 16/332 (2019.01)

(21) 国际申请号: PCT/CN2020/092963

(22) 国际申请日: 2020 年 5 月 28 日 (28.05.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010091180.4 2020年2月13日 (13.02.2020) 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: 陈秀玲 (CHEN, Xiuling); 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: 北京英特普罗知识产权代理有限公司 (INTELLECPRO CHINA LIMITED); 中国北京市西城区车公庄大街9号五栋大楼C座11层, Beijing 100044 (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,

WO 2021/159632 A1

(54) **Title:** INTELLIGENT QUESTIONING AND ANSWERING METHOD AND APPARATUS, COMPUTER DEVICE, AND COMPUTER STORAGE MEDIUM

(54) **发明名称:** 智能问答方法、装置、计算机设备及计算机存储介质

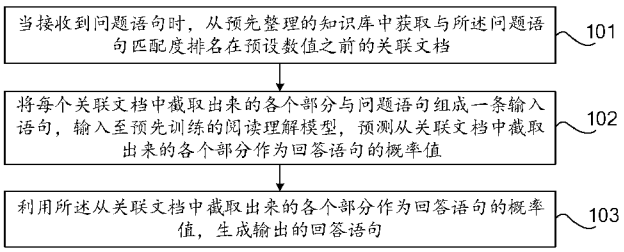


图 1

101 WHEN A QUESTION STATEMENT IS RECEIVED, OBTAIN ASSOCIATED DOCUMENTS HAVING MATCHING DEGREES RANKED IN FRONT OF A PRESET NUMERICAL VALUE WITH THE QUESTION STATEMENT FROM A PRE-ORGANIZED KNOWLEDGE BASE

102 COMBINE PARTS INTERCEPTED FROM EACH ASSOCIATED DOCUMENT AND THE QUESTION STATEMENT TO FORM AN INPUT STATEMENT, INPUT THE INPUT STATEMENT INTO A PRE-TRAINED READING COMPREHENSION MODEL, AND PREDICT PROBABILITY VALUES THAT THE PARTS INTERCEPTED FROM THE ASSOCIATED DOCUMENTS ARE USED AS AN ANSWER STATEMENT

103 GENERATE AN OUTPUT ANSWER STATEMENT BY USING THE PROBABILITY VALUES THAT THE PARTS INTERCEPTED FROM THE ASSOCIATED DOCUMENTS ARE USED AS AN ANSWER STATEMENT

(57) **Abstract:** An intelligent questioning and answering method and apparatus and a computer storage medium, relating to the technical field of artificial intelligence, and aiming to understand a question statement input by a user, analyze a search intention of the user, and provide a high-quality answer statement for the user. The method comprises: when a question statement is received, obtaining associated documents having matching degrees ranked in front of a preset numerical value with the question statement from a pre-organized knowledge base (101); combining parts intercepted from each associated document and the question statement to form an input statement, inputting the input statement into a pre-trained reading comprehension model, and predicting probability values that the parts intercepted from the associated documents are used as an answer statement (102); and generating an output answer statement by using the probability values that the parts intercepted from the associated documents are used as an answer statement (103).

(57) **摘要:** 一种智能问答方法、装置及计算机存储介质，涉及人工智能技术领域，能够对用户输入的问题语句进行理解，分析用户检索意图，给出用户优质的回答语句。所述方法包括：当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档（101）；将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值（102）；利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句（103）。

[见续页]

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

智能问答方法、装置、计算机设备及计算机存储介质

相关申请的交叉引用

本申请申明享有 2020 年 02 月 13 日递交的申请号为 CN202010091180.4、名称为“智能问答方法、装置、计算机设备及计算机存储介质”的中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本申请涉及人工智能技术领域，尤其是涉及到智能问答方法、装置、计算机设备及计算机存储介质。

背景技术

随着互联网的快速发展,基于关键词的传统搜索引擎由于准确率低、存在冗余信息以及需要用户对搜索结果进行甄别等缺陷,已不能很好地满足互联网信息检索需求。而问答系统允许用户使用自然语言问句进行信息查询，能够对用户输入的问句进行理解，分析用户检索意图，给出高质量的答案，不仅符合人们的检索习惯，而且提高了信息查询的效率。

开放域智能问答为问答系统的一种形式，主要采用基于问答对语料的相似度匹配，或者基于知识图谱检索，以及基于深度学习的生成式问答，能够实现基本语句的问答。然而，发明人意识到，在面对日新月异、种类繁多的开放域问答时，基于问答对语料的相似度匹配，或者基于知识图谱检索，需要维护一个大规模的问答对语料，或者知识图谱三元组语料，使得知识库经常出现覆盖不全、更新不及时而无法回答用户的问题，导致生成式问答的准确率并不能达到顺畅通话的要求，无法迅速给出较为准确的答案。

发明内容

有鉴于此，本申请提供了一种智能问答方法、装置、计算机设备及计算机存储介质，主要目的在于解决目前生成式问答的准确率不能达到顺畅通话的问题。

依据本申请一个方面，提供了一种智能问答方法，该方法包括：当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

依据本申请另一个方面，提供了一种智能问答装置，所述装置包括：获取单元，用于当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；预测单元，用于将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；生成单元，用于利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

依据本申请又一个方面，提供了一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现以下步骤：当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

依据本申请再一个方面，提供了一种计算机存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现以下步骤：当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；利用所述从关联文档中截取出来的各

个部分作为回答语句的概率值，生成输出的回答语句。

借由上述技术方案，本申请提供一种智能问答方法及装置，当接收到问题语句时，通过从预先整理的知识库中获取与问题语句匹配度排名在预设数值之前的关联文档，，进一步将关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值，从而生成输出的回答语句。与现有技术中智能问答方法相比，本申请预先整理的知识库中记录有从各个网站上整理的文档集合，提供了更完善的问答数据库，利用预先训练的阅读理解模型，能够针对用户输入的问题语句进行理解，分析用户检索意图，预测出关联文档中作为回答语句的概率值，给出用户优质的回答语句，提高了生成式问答的准确率。

附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本申请的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

图 1 示出了本申请实施例提供的一种智能问答方法的流程示意图；

图 2 示出了本申请实施例提供的另一种智能问答方法的流程示意图；

图 3 示出了本申请实施例提供的对问答语料进行情绪识别过程的示意图；

图 4 示出了本申请实施例提供的一种智能问答装置的结构示意图；

图 5 示出了本申请实施例提供的另一种智能问答装置的结构示意图。

具体实施方式

下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

本申请实施例提供了一种智能问答方法，能够对用户输入的问题语句进行理解，分析用户检索意图，给出用户优质的回答语句，如图 1 所示，该方法包括：

101、当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档。

其中，预先整理的知识库可以为维基百科知识库，维基百科知识库是一个类似于百度的网站，它会把网站内容整理成一个开放域的知识库，该知识库中包括各种文档以及数据集合，供各种人工智能算法训练时进行下载和使用。

可以理解的是，由于现有的问答领域中开放性问题太宽泛、话题所要求的时效性高，使得固定的语料库无法满足问答需求，本申请采用维基百科知识库作为问答语料，解决了在海量问题语句中搜索回答语句时兼顾搜索速度以及问题语句理解准确性的问题，给开放域智能问答提供一种可行性的思路。

具体在从预先整理的知识库中获取与问题语句匹配度排名在预设数值之前的关联文档时，由于预先整理的知识库中记录有从各个网站上整理的文档集合，这里可以根据问题语句在文档集合中各个文档中的重要程度，选取重要程度排名在预设数值之前的文档作为关联文档，当然还可以根据问题语句在文档集合中各个文档中出现的次数，选取次数排名在预设数值之前的文档作为关联文档，这里不进行限定。

102、将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值。

其中，预先训练的阅读理解模型通过使用 bert 预训练模型对问答数据集进行阅读理解任务的 fine-tune 训练和预测。这里使用 bert 预训练模型是一个使用双向 transformer 结构的语言模型，包括预训练在预训练阶段和阅读理解阶段，对两类任务进行训练，一类是把文档中 15%的词 mask 起来，在训练过程中对这些 mask 的词进行预测；另一类是预测一个句子对中的后面一个句子是否为前面一个句子的下一句话。

通过这两类任务的训练，生成问题语句以及关联文档中每个词的词向量、词向量的位置信息、词的语义信息，由于这些词向量包含自言语言文本的上下文语义信息，从关联文档中截取的各个部分都有可能是问题语句相应的回答语句，进一步根据问题语句以及关联文档中每个词的词向量、词向量的位置信息、词的语义信息，对从关联文档中截取出的各个部分进行预测，从而得到从关联文档中截取出来的各个部分作为回答语句的概率值。

103、利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

由于关联文本是采用维基百科作为问答语料，使用文本追踪后检索出的与问题语句关联程度较高的文档，所以通常都会记载有回答语句，通过阅读理解模型来预测关联文档中的回答语句，既解决了开放领域问题中话题要求时效性高而致使固定的语料库无法满足需要的特点，又解决了在海量问题中搜索答案时无法兼顾搜索速度和问题语义理解的准确性等问题，提高了问题回答的速度和准确性。

本申请实施例提供的一种智能问答方法，当接收到问题语句时，通过从预先整理的知识库中获取与问题语句匹配度排名在预设数值之前的关联文档，进一步将关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值，从而生成输出的回答语句。与现有技术中智能问答方法相比，本申请预先整理的知识库中记录有从各个网站上整理的文档集合，提供了更完善的问答数据库，利用预先训练的阅读理解模型，能够针对用户输入的问题语句进行理解，分析用户检索意图，预测出关联文档中作为回答语句的概率值，给出用户优质的回答语句，提高了生成式问答的准确率。

本申请实施例提供了另一种智能问答方法，能够对用户输入的问题语句进行理解，分析用户检索意图，给出用户优质的回答语句，如图 2 所示，所述方法包括：

201、对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引。

可以理解的是，对于预先整理的知识库中的文档，为了提高文档追踪效率，在系统初始化时首先对预先整理的知识库中文档集合中的各个文档进行分词处理，然后建立分词到文档集合中各个文档的倒排索引。

应说明的是，本申请对分词处理的方式不进行限定，可以使用分词工具如结巴分词、LTP、HanLP 等。

通过建立分词到文档集合中各个文档的倒排索引，对文档集合中各个文档以及分词进行编号，从而基于文档中包含的分词特征，能够迅速从海量的文档集合中查找出与问题语句相关的文档。例如，文档集合包含 5 个文档，通过对文档进行分词处理，得到文档所包含的各个分词，每个分词都有相应编号，同时记录分词出现所在的文档编号，分词 A 出现在文档 001、003 中，分词 B 出现在文档 004 中，分词 C 出现在文档 001、004 中，分词 D 出现在文档 005 中等等，相应的，分词 A 对应的倒排列表为{001、003}，分词 B 对应的倒排列表为{004}，分词 C 对应的倒排列表为{001、004}。

202、利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频。

其中，Bag-of-words 词袋模型最初被用在信息检索领域，对于一篇文档来说，假定不考虑文档内的词的顺序关系和语法，只考虑该文档是否出现过这个单词以及该单词出现的次数（词频）。这样一个文档的特征即表现为这个文档中所出现的单词以及每个单词出现的次数。

对于本申请实施例，具体在利用词袋模型可以对文档集合中各个文档的分词进行词频统计，得到各个分词在各个文档中出现的词频之后，还可以基于分词到文档集合中各个文档的倒排索引，将每个分词的词频在文档集合中各个文档出现的词频加入到分词的倒排列表中，例如，分词 A 对应的倒排列表为{001、003}，分词 A 在编号为 001 的文档中出现次

数为 1 词，在编号为 003 的文档中出现 4 次，相应的，分词 A 的倒排列表更新为{(001；1)、(003；4)}，得到分词在各个文档中的倒排列表，该倒排列表中记录有分词出现所在的各个文档以及在各个文档中的词频。

203、当接收到问题语句时，基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值。

具体在基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值的过程中，可以针对问题语句进行分词处理，得到问题语句所包含的分词，并基于建立的分词到文档集合中各个文档的倒排索引，获取对问题语句所包含的分词在各个文档中的词频，进一步计算问题语句所包含的分词在各个文档中的重要程度的评估值。

对于本申请实施例，计算问题语句所包含的分词在各个文档中的重要程度的评估值可以通过计算问题语句所包含的分词在各个文档中的 tf-idf 值，tf-idf 是一种用于信息检索与文本挖掘的常用加权技术。tf-idf 也是一种统计方法，用以评估一字词对于一个文件集或一个语料库的其中一份文件的重要程度。字词的重要性随着它在文件中出现的次数成正比，但同时会随着它在语料库中出现的频率成反比下降。

具体在基于分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值的过程中，倒排索引中记录有文档集合中文档数量、问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量，首先根据文档集合中文档数量、问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量，计算问题语句中各个分词在各个文档中的 tf-idf 值；然后汇总问题语句中各个分词在各个文档中的 tf-idf 值，得到问题语句在各个文档中的 tf-idf 值。

具体在计算问题语句中各个分词在各个文档中的 tf-idf 值过程中，首先获取问题语句中的分词在每个文档中出现的词频 termFreq，然后获取每个文档中出现的总的分词数 doctoTotalTerm；则$\mathrm{tf}=\sqrt{\mathrm{termFreq/docTotalTerm}}$；获取总的文档数为 docNum、包含问题语句中分词的文档数为 wordIndocNum，则 idf=1.0+log (docNum/(wordIndocNum+1))；问题语句中每个分词的 tf-idf=tf*idf；问题语句的 tf-idf 为该问题语句中所有分词的 tf-idf 之和/该问题语句的分词数量。

示例性的，问题语句划分为分词 A、分词 B、分词 C，分词 A 在文档 1 中出现的词频 termFreq=10，分词 A 在文档 2 中出现的词频 termFreq=7；分词 B 在文档 1 中出现的词频 termFreq=5，分词 B 在文档 2 中出现的词频 termFreq=0；分词 C 在文档 1 中出现的词频 termFreq=20，分词 C 在文档 2 中出现的词频 termFreq=10；文档 1 中的分词量为 100，文档 2 中的分词量为 140；则分词 A 在文档 1 中的$\mathrm{tf}=\sqrt{10/100}$，分词 A 在文档 2 中的$\mathrm{tf}=\sqrt{7/140}$；分词 B 在文档 1 中的$\mathrm{tf}=\sqrt{5/100}$，分词 B 在文档 2 中的$\mathrm{tf}=\sqrt{0/140}$；分词 C 在文档 1 中的$\mathrm{tf}=\sqrt{20/100}$，分词 C 在文档 2 中的$\mathrm{tf}=\sqrt{10/140}$；分词 A 的 idf=1.0+log (2/(2+1))；分词 B 的 idf=1.0+log (2/(2+1))；分词 C 的 idf=1.0+log (2/(1+1))；那么分词 A 在文档 1 中的 tf-idf 值=分词 A 在文档 1 中的$\mathrm{tf}=\sqrt{10/100}$*分词 A 的 idf；分词 A 在文档 2 中的 tf-idf 值=分词 A 在文档 2 中的$\mathrm{tf}=\sqrt{7/140}$*分词 A 的 idf；分词 B 在文档 1 中的 tf-idf 值=分词 B 在文档 1 中的$\mathrm{tf}=\sqrt{5/100}$*分词 B 的 idf；分词 B 在文档 2 中的 tf-idf 值=分词 B 在文档 2 中的$\mathrm{tf}=\sqrt{0/140}$*分词 B 的 idf；分词 C 在文档 1 中的 tf-idf 值=分词 C 在文档 1 中的$\mathrm{tf}=\sqrt{20/100}$*分词 C 的 idf；分词 C 在文档 2 中的 tf-idf 值=分词 C 在文档 2 中的$\mathrm{tf}=\sqrt{10/140}$*分词 C 的 idf，最终计算问题语句在文档 1 中的 tf-idf 值=分词 A 在文档 1 中的 tf-idf 值+分词 B 在文档 1 中的 tf-idf 值+分词 C 在文档 1 中的 tf-idf 值/3；计算问题语句在文档 2 中的 tf-idf 值=分词 A 在文档 2 中的 tf-idf 值+分词 B 在文档 2 中的 tf-idf 值+分词 C 在文档 2 中的 tf-idf 值/3。

204、按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档。

由于 tf-idf 值越大说明该文档对于问题语句越重要，与问题语句匹配度越高，所以评估值越大，说明问题语句与文档的关联程度越高，从而选取评估值排名在预设数值之前的文档作为关联文档。

可以理解的是，如果选取过多的关联文档后续识别工作量同样过多，影响智能对话的回答速度，这里优选的关联文档数量为 5 至 10。

205、在预训练阶段，将每个关联文档的部分词遮住，输入至预先训练的阅读理解模型对遮住的部分词进行预测，得到问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息。

对于本申请实施例，在预训练阶段，bert 预训练模型通过将每个关联文档中的部分词遮住，然后利用分词的上下文信息来预测这个分词原本的语义信息，使得学习到的语义信息可以融合一个分词左右两侧的上下文信息，进而提取问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息。

206、在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，输入至预先训练的阅读理解模型预测从关联文档中截取出来的各个部分作为回答语句的概率值。

对于本申请实施例，在阅读理解阶段，具体通过对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，得到词编码和位置编码，将词编码与位置编码之间的运算结果，输入至预先训练的阅读理解模型以使得位置信息补充到词编码之中，获取问题语句与从关联文档中截取出来的各个部分之间的关联关系，基于问题语句与从关联文档中截取出来的各个部分之间的关联关系，预测从关联文档中截取出来的各个部分作为回答语句的概率值。

具体在问题语句以及关联文档中，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码的过程，词向量为每个分词对应一个 768 维度的向量；位置信息为预先给每个分词标记的一个整数位，后续也根据这个整数位转换为一个 768 维度的向量；语义信息为在阅读理解模型中将问题语句与关联文档区分开，所有问题语句中的分词标注为 0，所有关联文档中的分词标注为 1，后续将 0 和 1 转换为一个 768 维度的向量，阅读理解阶段中预先训练的阅读理解模型的输入即为词向量、位置向量以及语义向量的相加，问题语句对应的回答语句为从关联文档中截取的一段文本，假设回答语句在关联文档中的开始位置 start-point，回答语句在关联文档中的结束位置 end-point，通过预先训练的阅读理解模型可以预测出从文档中截取出来的各个部分中每个分词作为 start-point 和 end-point 的概率值。

207、根据筛选指令对从关联文档中截取出来的各个部分作为回答语句的概率值进行排序。

可以理解的是，通常情况下，从关联文档中截取出来的各个部分作为回答语句的概率值越高，说明该部分的内容更适合作为回答语句，可以选取截取出来作为回答语句概率值最高的部分文档生成输出的回答语句，从而向用户提供更准确的回答内容。

进一步地，为了提高输出回答语句的灵活性，用户在输入问题语句的时候，可能考虑的话题实际应用场合以及语境等场景因素，从关联文档中截取出来的各个部分作为回答语句的概率值最高的文档并非满足场景因素，所以在生成输出的回答语句之前，可以通过设置筛选指令来结合用户当前的场景因素对从关联文档中截取出来的各个部分作为回答语句的概率值进行排序，从而选取更适合的部分文档生成输出的回答语句，这里对场景因素不进行限定。

208、获取从关联文档中截取出来的各个部分作为回答语句概率值最高的部分文档，生成回答语句。

对于本申请实施例，具体智能问答的流程可以如图 3 所示，当输入用户的问题语句时，通过从维基百科知识库中实时追踪与问题语句相关联的文档，选取排名前 5 的关联文档，将问题语句与关联文档中的句子输入至预先训练的阅读理解模型进行短文本阅读理解，从而将预测得到文档中的句子作为回答语句的概率值，并选取概率值靠前的回答句子作为最佳答案。

进一步地，作为图 1 所述方法的具体实现，本申请实施例提供了一种智能问答装置，如图 4 所示，所述装置包括：获取单元 31、预测单元 32、生成单元 33。

获取单元 31，可以用于当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；

预测单元 32，可以用于将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；

生成单元 33，可以用于利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

本申请实施例提供的一种智能问答装置，当接收到问题语句时，通过从预先整理的知识库中获取与问题语句匹配度排名在预设数值之前的关联文档，，进一步将关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值，从而生成输出的回答语句。与现有技术中智能问答方法相比，本申请预先整理的知识库中记录有从各个网站上整理的文档集合，提供了更完善的问答数据库，利用预先训练的阅读理解模型，能够针对用户输入的问题语句进行理解，分析用户检索意图，预测出关联文档中作为回答语句的概率值，给出用户优质的回答语句，提高了生成式问答的准确率。

作为图 4 中所示智能问答装置的进一步说明，图 5 是根据本申请实施例另一种智能问答装置的结构示意图，如图 5 所示，所述预先整理的知识库中记录有从各个网站上整理的文档集合，所述装置还包括：

建立单元 34，可以用于在所述当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档之前对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引；

统计单元 35,可以用于利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频。

在具体应用场景中，如图 5 所示，所述获取单元 31 包括：

计算模块 311，可以用于基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值；

选取模块 312，可以用于按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档。

在具体应用场景中，所述计算模块 311，具体可以用于对所述问题语句进行分词处理，基于所述分词到文档集合中各个文档的倒排索引，查询问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量；

所述计算模块 311，具体还可以用于根据文档集合中文档数量、问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量，计算问题语句中各个分词在各个文档中重要程度的评估值；

所述计算模块 311，具体还可以用于汇总所述问题语句中各个分词在各个文档中重要程度的评估值，得到问题语句在各个文档中重要程度的评估值。

在具体应用场景中，如图 5 所示，所述预先训练的阅读理解模型通过使用 bert 预训练模型对问答数据集进行阅读理解任务的 fine-tune 训练和预测，包括预训练阶段和阅读理解阶段，所述预测单元 32 包括：

第一预测模块 321，可以用于在预训练阶段，将每个关联文档的部分词遮住，输入至预先训练的阅读理解模型对遮住的部分词进行预测，得到问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息；

第二预测模块 322，可以用于在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，输入至预先训练的阅读理解模型预测从关联文档中截取出来的各个部分作为回答语句的概率值。

进一步地，所述第二预测模块 322，具体可以用于在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，得到词编码和位置编码；

所述第二预测模块 322，具体还可以用于将词编码与位置编码之间的运算，输入至预先训练的阅读理解模型以使得位置信息补充到词编码之中，获取问题语句与从关联文档中截取出来的各个部分之间的关联关系；

所述第二预测模块 322，具体还可以用于基于所述问题语句与从关联文档中截取出来的各个部分之间的关联关系，预测从关联文档中截取出来的各个部分作为回答语句的概率值。

在具体应用场景中，如图 5 所示，所述生成单元 33 包括：

排序模块 331，可以用于根据筛选指令对从关联文档中截取出来的各个部分作为回答语句的概率值进行排序；

生成模块 332，可以用于获取从关联文档中截取出来的各个部分作为回答语句概率值最高的部分文档，生成回答语句。

需要说明的是，本实施例提供的一种智能问答装置所涉及各功能单元的其他相应描述，可以参考图 1-图 2 中的对应描述，在此不再赘述。

基于上述如图 1、图 2 所示方法，相应的，本实施例还提供了一种存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述如图 1、图 2 所示的智能问答方法。

基于这样的理解，本申请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易失性存储介质（例如静态 RAM 和动态内存 DRAM 等）中，也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，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施场景所述的方法。

基于上述如图 1、图 2 所示的方法，以及图 4、图 5 所示的虚拟装置实施例，为了实现上述目的，本申请实施例还提供了一种计算机设备，具体可以为个人计算机、服务器、网络设备等，该实体设备包括存储介质和处理器；存储介质，用于存储计算机程序；处理器，用于执行计算机程序以实现上述如图 1、图 2 所示的智能问答方法。

可选地，该计算机设备还可以包括用户接口、网络接口、摄像头、射频（Radio Frequency，RF）电路，传感器、音频电路、WI-FI 模块等等。用户接口可以包括显示屏（Display）、输入单元比如键盘（Keyboard）等，可选用户接口还可以包括 USB 接口、读卡器接口等。网络接口可选的可以包括标准的有线接口、无线接口（如蓝牙接口、WI-FI 接口）等。

本领域技术人员可以理解，本实施例提供的智能问答装置的实体设备结构并不构成对该实体设备的限定，可以包括更多或更少的部件，或者组合某些部件，或者不同的部件布置。

存储介质中还可以包括操作系统、网络通信模块。操作系统是管理上述计算机设备硬件和软件资源的程序，支持信息处理程序以及其它软件和/或程序的运行。网络通信模块用于实现存储介质内部各组件之间的通信，以及与该实体设备中其它硬件和软件之间通信。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到本申请可以借助软件加必要的通用硬件平台的方式来实现，也可以通过硬件实现。通过应用本申请的技术方案，与目前现有技术相比，本申请预先整理的知识库中记录有从各个网站上整理的文档集

合，提供了更完善的问答数据库，利用预先训练的阅读理解模型，能够针对用户输入的问题语句进行理解，分析用户检索意图，预测出关联文档中作为回答语句的概率值，给出用户优质的回答语句，提高了生成式问答的准确率。

本领域技术人员可以理解附图只是一个优选实施场景的示意图，附图中的模块或流程并不一定是实施本申请所必须的。本领域技术人员可以理解实施场景中的装置中的模块可以按照实施场景描述进行分布于实施场景的装置中，也可以进行相应变化位于不同于本实施场景的一个或多个装置中。上述实施场景的模块可以合并为一个模块，也可以进一步拆分成多个子模块。

上述本申请序号仅仅为了描述，不代表实施场景的优劣。以上公开的仅为本申请的几个具体实施场景，但是，本申请并非局限于此，任何本领域的技术人员能思之的变化都应落入本申请的保护范围。

权利要求书

1、一种智能问答方法，其中，所述方法包括：

当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；

将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；

利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

2、根据权利要求 1 所述的智能问答方法，其中，所述预先整理的知识库中记录有从各个网站上整理的文档集合，在所述当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档之前，所述方法还包括：

对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引；

利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频。

3、根据权利要求 2 所述的智能问答方法，其中，所述从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档，具体包括：

基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值；

按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档。

4、根据权利要求 3 所述的智能问答方法，其中，所述基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值，具体包括：

对所述问题语句进行分词处理，基于所述分词到文档集合中各个文档的倒排索引，查询问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量；

根据文档集合中文档数量、问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量，计算问题语句中各个分词在各个文档中重要程度的评估值；

汇总所述问题语句中各个分词在各个文档中重要程度的评估值，得到问题语句在各个文档中重要程度的评估值。

5、根据权利要求 1 所述的智能问答方法，其中，所述预先训练的阅读理解模型通过使用 bert 预训练模型对问答数据集进行阅读理解任务的 fine-tune 训练和预测，包括预训练阶段和阅读理解阶段，所述将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值，具体包括：

在预训练阶段，将每个关联文档的部分词遮住，输入至预先训练的阅读理解模型对遮住的部分词进行预测，得到问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息；

在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，输入至预先训练的阅读理解模型预测从关联文档中截取出来的各个部分作为回答语句的概率值。

6、根据权利要求 5 所述的智能问答方法，其中，所述在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，输入至预先训练的阅读理解模型预测从关联文档中截取出来的各个部分作为回答语句的概率值，具体包括：

在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，得到词编码和位置编码；

将词编码与位置编码之间的运算结果，输入至预先训练的阅读理解模型以使得位置信息补充到词编码之中，获取问题语句与从关联文档中截取出来的各个部分之间的关联关系；

基于所述问题语句与从关联文档中截取出来的各个部分之间的关联关系，预测从关联文档中截取出来的各个部分作为回答语句的概率值。

7、根据权利要求 1 所述的智能问答方法，其中，所述利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句，具体包括：

根据筛选指令对从关联文档中截取出来的各个部分作为回答语句的概率值进行排序；

获取从关联文档中截取出来的各个部分作为回答语句概率值最高的部分文档，生成回答语句。

8、一种智能问答装置，其中，所述装置包括：

获取单元，用于当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；

预测单元，用于将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；

生成单元，用于利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

9、根据权利要求 8 所述的智能问答装置，其中，所述预先整理的知识库中记录有从各个网站上整理的文档集合，所述智能问答装置还包括：

建立单元，用于在所述当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档之前，对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引；

统计单元，用于利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频。

10、根据权利要求 9 所述的智能问答装置，其中，所述获取单元包括：

计算模块，用于基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值；

选取模块，用于按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档。

11、根据权利要求 10 所述的智能问答装置，其中，所述计算模块具体用于对所述问题语句进行分词处理，基于所述分词到文档集合中各个文档的倒排索引，查询问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量；

所述计算模块具体还用于根据文档集合中文档数量、问题语句中各个分词在文档集合中各个文档出现的词频、各个文档中包含的分词量，计算问题语句中各个分词在各个文档中重要程度的评估值；

所述计算模块具体还用于汇总所述问题语句中各个分词在各个文档中重要程度的评估值，得到问题语句在各个文档中重要程度的评估值。

12、根据权利要求 8 所述的智能问答装置，其中，所述预先训练的阅读理解模型通过使用 bert 预训练模型对问答数据集进行阅读理解任务的 fine-tune 训练和预测，包括预训练阶段和阅读理解阶段，所述预测单元包括：

第一预测模块，用于在预训练阶段，将每个关联文档的部分词遮住，输入至预先训练的阅读理解模型对遮住的部分词进行预测，得到问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息；

第二预测模块，用于在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，输入至预先训练的阅读理解模型预测从关联文档中截取出来的各个部分作为回答语句的概率值。

13、根据权利要求 12 所述的智能问答装置，其中，所述第二预测模块具体用于在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，得到词编码和位置编码；

所述第二预测模块具体还用于将词编码与位置编码之间的运算结果，输入至预先训练的阅读理解模型以使得位置信息补充到词编码之中，获取问题语句与从关联文档中截取出来的各个部分之间的关联关系；

所述第二预测模块具体还用于基于所述问题语句与从关联文档中截取出来的各个部分之间的关联关系，预测从关联文档中截取出来的各个部分作为回答语句的概率值。

14、根据权利要求 8 所述的智能问答装置，其中，所述生成单元包括：

排序模块，用于根据筛选指令对从关联文档中截取出来的各个部分作为回答语句的概率值进行排序；

生成模块，用于获取从关联文档中截取出来的各个部分作为回答语句概率值最高的部分文档，生成回答语句。

15、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其中，所述处理器执行所述计算机程序时实现以下步骤：

当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；

将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；

利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

16、根据权利要求 15 所述的计算机设备，其中，所述处理器执行所述计算机程序时还实现以下步骤：

所述预先整理的知识库中记录有从各个网站上整理的文档集合，在所述当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档之前，对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引；利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频。

17、根据权利要求 16 所述的计算机设备，其中，所述从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档，具体包括：

基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值；

按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档。

18、一种计算机存储介质，其上存储有计算机程序，其中，所述计算机程序被处理器执行时实现以下步骤：

当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档；

将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值；

利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句。

19、根据权利要求 18 所述的计算机存储介质，其中，所述计算机程序被处理器执行时还实现以下步骤：

所述预先整理的知识库中记录有从各个网站上整理的文档集合，在所述当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文

档之前，对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引；利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频。

20、根据权利要求 19 所述的计算机存储介质，其中，所述从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档，具体包括：

基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值；

按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档。

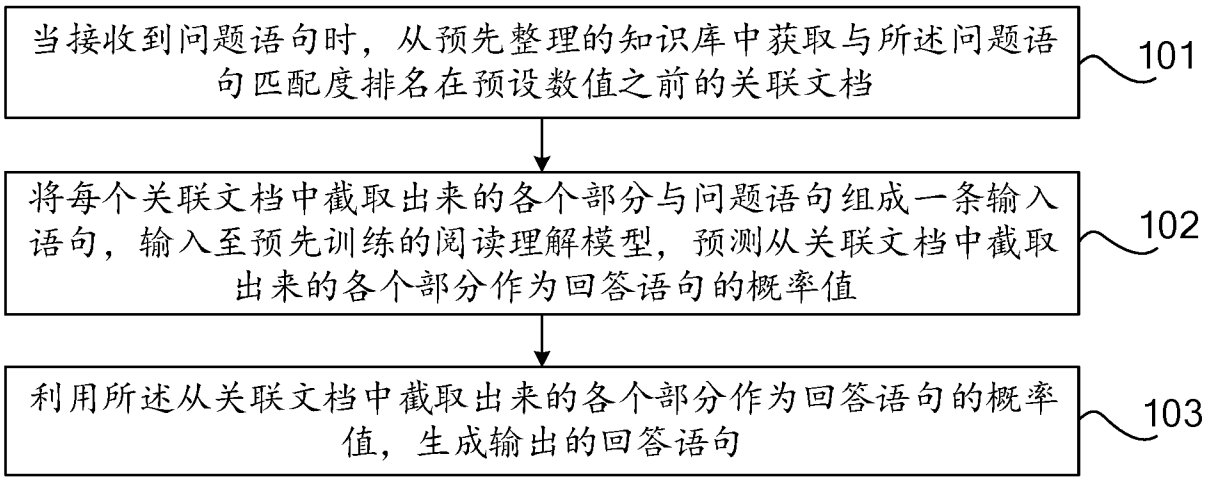

当接收到问题语句时，从预先整理的知识库中获取与所述问题语句匹配度排名在预设数值之前的关联文档
101
将每个关联文档中截取出来的各个部分与问题语句组成一条输入语句，输入至预先训练的阅读理解模型，预测从关联文档中截取出来的各个部分作为回答语句的概率值
102
利用所述从关联文档中截取出来的各个部分作为回答语句的概率值，生成输出的回答语句
103

图 1

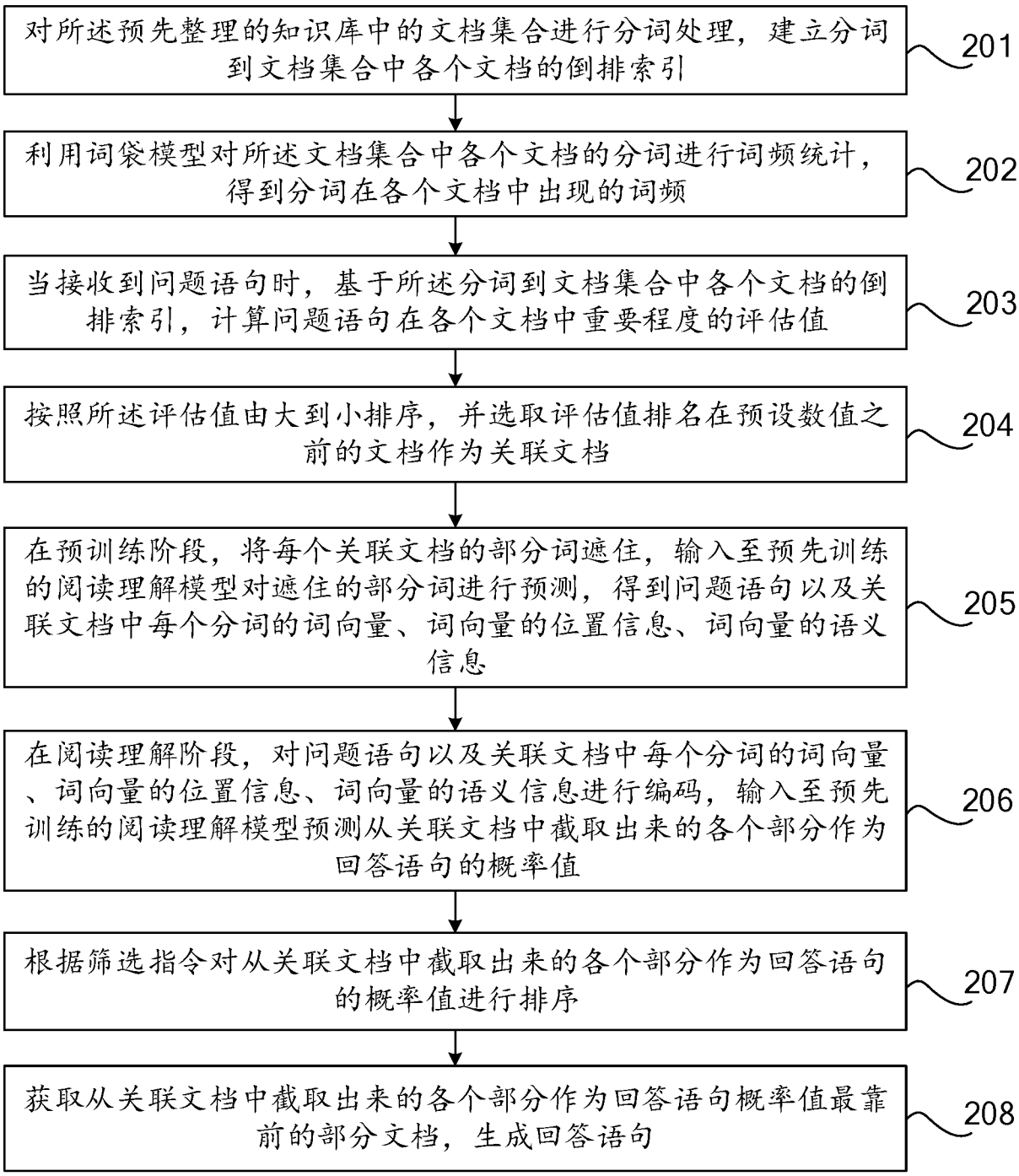

对所述预先整理的知识库中的文档集合进行分词处理，建立分词到文档集合中各个文档的倒排索引
201
利用词袋模型对所述文档集合中各个文档的分词进行词频统计，得到分词在各个文档中出现的词频
202
当接收到问题语句时，基于所述分词到文档集合中各个文档的倒排索引，计算问题语句在各个文档中重要程度的评估值
203
按照所述评估值由大到小排序，并选取评估值排名在预设数值之前的文档作为关联文档
204
在预训练阶段，将每个关联文档的部分词遮住，输入至预先训练的阅读理解模型对遮住的部分词进行预测，得到问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息
205
在阅读理解阶段，对问题语句以及关联文档中每个分词的词向量、词向量的位置信息、词向量的语义信息进行编码，输入至预先训练的阅读理解模型预测从关联文档中截取出来的各个部分作为回答语句的概率值
206
根据筛选指令对从关联文档中截取出来的各个部分作为回答语句的概率值进行排序
207
获取从关联文档中截取出来的各个部分作为回答语句概率值最靠前的部分文档，生成回答语句
208

图 2

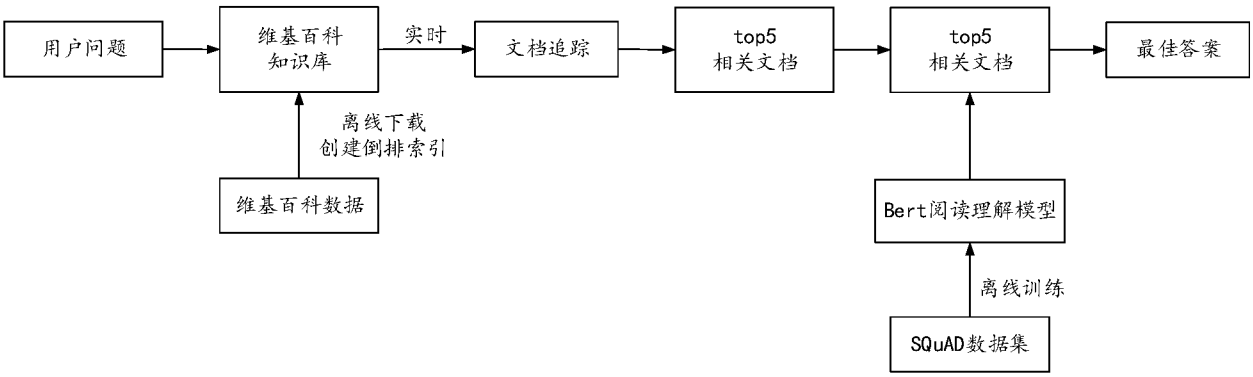

用户问题
维基百科
知识库
实时
文档追踪
top5
相关文档
top5
相关文档
最佳答案
离线下载
创建倒排索引
维基百科数据
Bert阅读理解模型
离线训练
SQuAD数据集

图 3

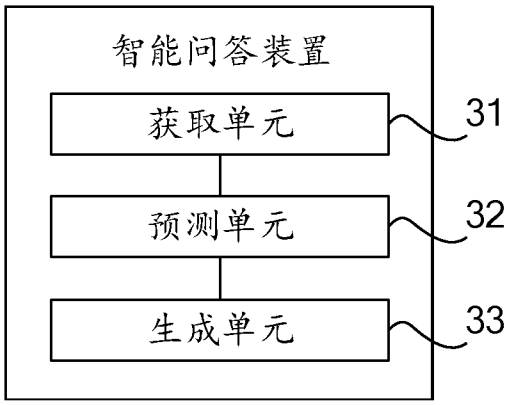

智能问答装置
获取单元
31
预测单元
32
生成单元
33

图 4

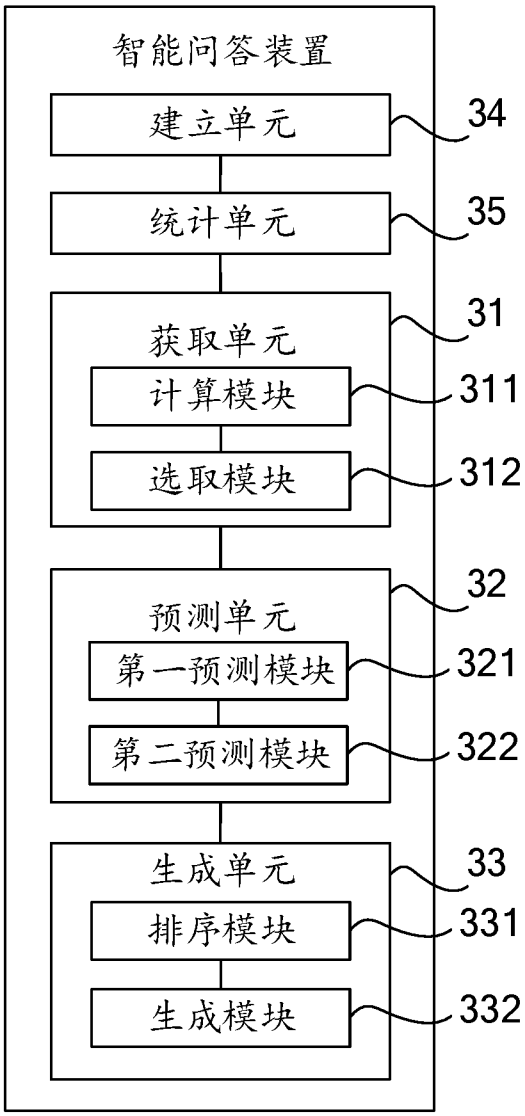

智能问答装置
建立单元
34
统计单元
35
获取单元
31
计算模块
311
选取模块
312
预测单元
32
第一预测模块
321
第二预测模块
322
生成单元
33
排序模块
331
生成模块
332

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/092963

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/332(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)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT; CNKI; WPI; EPODOC: 智能, 问答, 问题, 语句, 意图, 回答, 匹配, 关联, 训练, 模型, 概率, 截取, 分词, 评估, 阅读, 理解, bert, intelligence, question, answer, sentence, intent, match, association, training, model, probability, interception, word, segment, evaluate, read, understand

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 110390003 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 29 October 2019 (2019-10-29) description, paragraphs [0026]-[0118]	1-20
A	CN 110502621 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 26 November 2019 (2019-11-26) entire document	1-20
A	CN 108446320 A (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 24 August 2018 (2018-08-24) entire document	1-20
A	US 2018276525 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 27 September 2018 (2018-09-27) entire document	1-20
A	CN 110309283 A (ALIBABA GROUP HOLDING LIMITED) 08 October 2019 (2019-10-08) 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	Date of mailing of the international search report
02 November 2020	**18 November 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2020/092963

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 110390003 A	29 October 2019	None	
CN 110502621 A	26 November 2019	None	
CN 108446320 A	24 August 2018	None	
US 2018276525 A1	27 September 2018	CN 106844368 A	13 June 2017
		WO 2017092380 A1	08 June 2017
		CN 105865321 A	17 August 2016
CN 110309283 A	08 October 2019	None	

国际检索报告

国际申请号
PCT/CN2020/092963

A. 主题的分类

G06F 16/332(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)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT;CNKI;WPI;EPODOC:智能，问答，问题，语句，意图，回答，匹配，关联，训练，模型，概率，截取，分词，评估，阅读，理解，bert，intelligence，question，answer，sentence，intent，match，association，training，model，probability，interception，word，segment，evaluate，read，understand

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 110390003 A (北京百度网讯科技有限公司) 2019年 10月 29日 (2019-10-29) 说明书第[0026]-[0118]段	1-20
A	CN 110502621 A (平安科技深圳有限公司) 2019年 11月 26日 (2019-11-26) 全文	1-20
A	CN 108446320 A (北京搜狗科技发展有限公司) 2018年 8月 24日 (2018-08-24) 全文	1-20
A	US 2018276525 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 2018年 9月 27日 (2018-09-27) 全文	1-20
A	CN 110309283 A (阿里巴巴集团控股有限公司) 2019年 10月 8日 (2019-10-08) 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	国际检索报告邮寄日期
2020年 11月 2日	2020年 11月 18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 王璐 电话号码 86-(10)-53961303

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2020/092963

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	110390003	A	2019年 10月 29日	无			
CN	110502621	A	2019年 11月 26日	无			
CN	108446320	A	2018年 8月 24日	无			
US	2018276525	A1	2018年 9月 27日	CN	106844368	A	2017年 6月 13日
				WO	2017092380	A1	2017年 6月 8日
				CN	105865321	A	2016年 8月 17日
CN	110309283	A	2019年 10月 8日	无			