

(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
2022 年 4 月 28 日 (28.04.2022)



(10) 国际公布号
WO 2022/083103 A1

(51) 国际专利分类号:
G06F 16/332 (2019.01)

(21) 国际申请号: PCT/CN2021/091700

(22) 国际申请日: 2021 年 4 月 30 日 (30.04.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202011149730.X 2020 年 10 月 23 日 (23.10.2020) 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: 刘东煜 (LIU, Dongyu); 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)。 陈乐清 (CHEN, Leqing); 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)。 曾增烽 (ZENG, Zengfeng); 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: 深圳市明日今典知识产权代理事务所 (普通合伙) (SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL)); 中国广东省深圳市南山区南头街道安乐社区关口二路 15 号智恒产业园 30 栋 4 层 406 室, Guangdong 518000 (CN)。

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

(54) 发明名称: 问题语句的改写方法、装置、计算机设备和存储介质

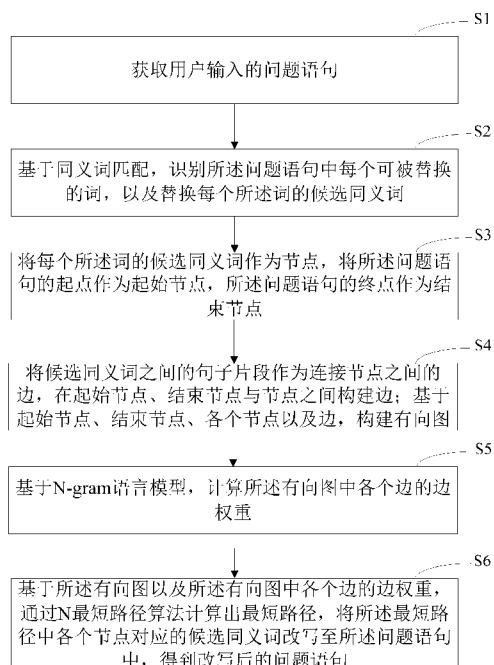


图 1

S1 Acquire a question statement inputted by a user
S2 On the basis of synonym matching, identify each replaceable word in the question statement and replace candidate synonyms for each word
S3 Use the candidate synonyms for each word as nodes, and set the start point of the question statement as a start node and the end point of the question statement as an end node
S4 Set sentence fragments between the candidate synonyms as edges between connecting nodes to build edges between the start node, the end node, and the nodes and, on the basis of the edges between the start node, the end node, and the nodes, build a directed graph
S5 On the basis of an n-gram language model, calculate an edge weighting of each edge in the directed graph
S6 On the basis of the directed graph and the edge weighting of each edge in the directed graph, and by means of an N shortest path algorithm, calculate the shortest path, and rewrite the candidate synonyms corresponding to the nodes in the shortest path to the question sentence to obtain a rewritten question sentence

(57) **Abstract:** The present application relates to the technical field of artificial intelligence, and provides a method and apparatus for rewriting question statements, a computer device, and a storage medium: acquiring a question statement inputted by a user; on the basis of synonym matching, identifying each replaceable word in the question statement and replacing candidate synonyms for each word; using the candidate synonyms for each word as nodes, and setting the start point of the question statement as a start node and the end point of the question statement as an end node; setting sentence fragments between the candidate synonyms as edges between

(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, IT, 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, 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))。

connecting nodes to build a directed graph; on the basis of an n-gram language model, calculating an edge weighting of each edge in the directed graph; and, by means of an N shortest path algorithm, calculating the shortest path, and rewriting the candidate synonyms corresponding to the nodes in the shortest path to the question sentence to obtain a rewritten question sentence. By means of a shortest path calculation, the present application finds the synonyms best conforming to the context, and thereby rewrites question statements.

(57) 摘要: 本申请涉及人工智能技术领域, 提供一种问题语句的改写方法、装置、计算机设备和存储介质, 获取用户输入的问题语句; 基于同义词匹配, 识别所述问题语句中每个可被替换的词, 以及替换每个所述词的候选同义词; 将每个所述词的候选同义词作为节点, 将所述问题语句的起点作为起始节点, 所述问题语句的终点作为结束节点; 将所述候选同义词之间的句子片段作为连接节点之间的边, 构建有向图; 基于N-gram语言模型, 计算所述有向图中各个边的边权重; 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句。本申请通过最短路径的计算, 找出最符合语境同义词, 从而对文图语句进行改写。

问题语句的改写方法、装置、计算机设备和存储介质

本申请要求于2020年10月23日提交中国专利局、申请号为202011149730X,发明名称为“问题语句的改写方法、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

本申请涉及人工智能技术领域,特别涉及一种问题语句的改写方法、装置、计算机设备和存储介质。

背景技术

在FAQ(问题解答)模块中,针对用户的问题,其回复答案生成主要将用户的问题与问题模板匹配的方式完成。但是在某些情况下用户问题与模板问题语义相近,却无法召回答案。比如:用户问“抑郁症可以投保吗?”,而模板问为“精神病可以投保吗?”,这种情况下可以通过同义词改写的方式扩充输入的问题语句,如:“抑郁症”的同义词为“精神病”,则可以在一定程度上提高召回率。

在具体的应用场景,比如在FAQ问答机器人中,数据库会事先存储多个问题的模板,当用户提问时会匹配到相应的模板问题从而进行自动回复。

然而,对于同一类问题,用户的提问方式却多种多样,同时也包含大量的口语表述,如:“注射疫苗->打疫苗”、“腿骨折->腿瘸了”、“取消->退保”等。

因此,在实际中虽然用户问题与模板问题的语义相近,却无法正确匹配到相应的回复。

因此,可以对用户的问题语句中的词进行同义词替换,但是发明人意识到在不同语境下同义词的合理性却存在差异。如“打疫苗”可以改写为“注射疫苗”,此时,“打”与“注射”是一组同义词,然而将“打篮球”改写为“注射篮球”却十分不合理。这样的例子普遍存在,如“可以投保吗?”可以改写为“能投保吗?”,而将“这个想法可以”改写为“这个想法能”却很滑稽。因此,目前的同义词替换存在不符合语境的缺陷。

技术问题

本申请的主要目的为提供一种问题语句的改写方法、装置、计算机设备和存储介质,旨在克服目前同义词改写时出现不符合语境的缺陷。

技术方案

为实现上述目的,本申请提供了一种问题语句的改写方法,包括以下步骤:

获取用户输入的问题语句;

基于同义词匹配,识别所述问题语句中每个可被替换的词,以及替换每个所述词的候选同义词;

将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，
5 构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

基于N-gram语言模型，计算所述有向图中各个边的边权重；

10 基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

本申请还提供了一种问题语句的改写装置，包括：

获取单元，用于获取用户输入的问题语句；

15 第一识别单元，用于基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；

第一构建单元，用于将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

20 第二构建单元，用于将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

25 第一计算单元，用于基于N-gram语言模型，计算所述有向图中各个边的边权重；

改写单元，用于基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

30 本申请还提供一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机程序，所述处理器执行所述计算机程序时实现一种问题语句的改写方法，包括：

获取用户输入的问题语句；

基于同义词匹配, 识别所述问题语句中每个可被替换的词, 以及替换每个所述词的候选同义词;

将每个所述词的候选同义词作为节点, 将所述问题语句的起点作为起始节点, 所述问题语句的终点作为结束节点;

- 5 将所述候选同义词之间的句子片段作为连接节点之间的边, 分别在所述起始节点、结束节点与所述节点之间构建边; 基于所述起始节点、结束节点、各个所述节点以及所述边, 构建有向图; 其中, 所述有向图中的各个节点对应的候选同义词之间不存在交叉, 所述有向图中由起始节点开始, 由结束节点结束, 且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点;
- 10 节点;

基于N-gram语言模型, 计算所述有向图中各个边的边权重;

基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句。

- 15 本申请还提供一种计算机可读存储介质, 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一种问题语句的改写方法, 包括:

获取用户输入的问题语句;

基于同义词匹配, 识别所述问题语句中每个可被替换的词, 以及替换每个所述词的候选同义词;

- 20 将每个所述词的候选同义词作为节点, 将所述问题语句的起点作为起始节点, 所述问题语句的终点作为结束节点;

- 将所述候选同义词之间的句子片段作为连接节点之间的边, 分别在所述起始节点、结束节点与所述节点之间构建边; 基于所述起始节点、结束节点、各个所述节点以及所述边, 构建有向图; 其中, 所述有向图中的各个节点对应的候选同义词之间不存在交叉, 所述有向图中由起始节点开始, 由结束节点结束, 且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点;
- 25 节点;

基于N-gram语言模型, 计算所述有向图中各个边的边权重;

- 基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句。
- 30

有益效果

本申请提供的问题语句的改写方法、装置、计算机设备和存储介质，获取用户输入的问题语句；基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；基于N-gram语言模型，计算所述有向图中各个边的边权重；通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。本申请通过最短路径的计算，找出最符合语境同义词，从而对文图语句进行改写。

附图说明

图1 是本申请一实施例中问题语句的改写方法步骤示意图；

图2 是本申请一实施例中问题语句的改写装置结构框图；

图3 为本申请一实施例的计算机设备的结构示意图。

本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

本发明的最佳实施方式

为了使本申请的目的、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。

参照图1，本申请一实施例中提供了一种问题语句的改写方法，包括以下步骤：

步骤S1，获取用户输入的问题语句；

步骤S2，基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；

步骤S3，将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

步骤S4，将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

步骤S5, 基于N-gram语言模型, 计算所述有向图中各个边的边权重;

步骤S6, 基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句。

5 在本实施例中, 上述方法可应用于智慧城市技术领域中, 以推动智慧城市的建设。具体地, 如上述步骤S1所述的, 应用于智能解答的场景中, 上述问题语句为用户在用户终端上语音输入或者文字输入的语句, 例如: 请问交费期间可以退保金鑫盛吗?

如上述步骤S2所述的, 配置有离线同义词挖掘系统, 该系统经过长期的同义词挖掘, 从而不断补充同义词词典。上述同义词词典中的同义词以key-value形式存储, key为待匹配的词语, value为若干可能与该词成为同义词的词语。基于上述同义词词典, 利用同义词匹
10 配, 从上述问题语句中识别出每个可被替换的词, 即识别出问题语句中每个具有同义词的词, 以及根据同义词词典获取到替换每个词的候选同义词。

如上述步骤S3-S4所述的, 将每个候选同义词看作为一个“节点”, 候选同义词与候选同义词之间的句子片段看作“边”, 构建有向图 $G(V, E)$ 。候选同义词 v_1 与候选同义词 v_2 不存在交叉则建立有向边 e_{12} , 设: 候选同义词 v_1 在句中的位置在候选同义词 v_2 之前, 则边 e_{12}
15 的方向由 v_1 指向 v_2 。同时添加起始节点“<Start>”与结束节点“<End>”, 上述起始节点以及结束节点与上述候选同义词对应的节点之间也依次构建边。如下图所示, 因“交费”、“期间”两个匹配词间无交叉, 所以存在有向边, 而因“交费期间”、“期间”两个匹配词间存在交叉则不存在有向边。

20 如上述步骤S5所述的, 上述N-gram语言模型用于计算上个各个候选同义词代入至问题语句后的句子通顺度分数, 并基于上述通顺度分数计算所述有向图中各个边的边权重。上述N-gram语言模型采用大规模通用语料(人民日报、搜狗语料等)以及垂直领域语料(如: FAQ中的语料)训练得到。其中, N-gram是一种统计语言模型算法, 通过对语句的出现概率进行建模从而得到句子困惑度分数(即顺畅度分数)。

25 如上述步骤S6所述的, 根据上述构造好的有向图, 使用N最短路径进行求解出起始节点至结束节点的最短路径(即概率分数最大), 针对最短路径, 依次选择路径上节点对应的候选同义词进行改写即得到了改写后的问题句子。

在一实施例中, 所述基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题
30 语句中, 得到改写后的问题语句的步骤S6之后, 包括:

步骤S7, 计算所述问题语句与所述改写后的问题语句的语义相似度;

步骤S8, 判断所述语义相似度是否大于阈值;

步骤S9, 若不大于, 则基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出第二短路径, 将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句;

5 步骤S10, 若大于, 则在数据库中匹配对应所述改写后的问题语句的答案, 并将匹配出的所述答案推送至所述用户所在的用户终端。

在本实施例中, 得到改写后问题语句, 为了避免上述模型训练不够造成改写不正确的缺陷, 还需要计算上述改写后的问题语句与用户原输入的问题语句的语义相似度; 当该语义相似度大于预设的阈值时, 则表明其语义与原问题语句的语义近似, 则可以在数据库中
10 获取该改写后的问题语句所对应的答案。若语义相似度不大于预设的阈值时, 则表明其语义与原问题语句的语义不够近似, 此时有可能是计算出现误差, 或者模型训练过程中出现问题。因此, 可以基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出第二短路径, 将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句。

15 在一实施例中, 所述计算所述问题语句与所述改写后的问题语句的语义相似度的步骤S7, 包括:

基于预设的编码器, 得到所述问题语句对应的第一向量矩阵, 以及得到所述改写后的问题语句对应的第二向量矩阵;

计算第一向量矩阵与所述第二向量矩阵的相似度。

20 相似度的计算公式为:

$$\text{Sim} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}};$$

其中, A_i 为所述第一向量矩阵的第 i 个分向量的数值, B_i 为所述第二向量矩阵的第 i 个分向量的数值, 所述第一向量矩阵和所述第二向量矩阵均包括 n 个分向量。

25 在一实施例中, 所述基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句的步骤S6之后, 包括:

步骤S71, 基于所述改写后的问题语句, 识别所述用户的意图;

步骤S81, 判断所述意图是否为指定意图; 所述指定意图为购买产品意图;

30 步骤S91, 若是, 则获取对应所述指定意图的产品信息文本, 并对所述产品信息文本进行哈希计算, 得到对应的哈希值存储于数据库中;

步骤S101, 将所述产品信息文本以及所述哈希值发送至所述用户所在的用户终端, 并提醒所述用户在所述产品信息文本中输入用户信息, 并基于所述哈希值对所述产品信息文本进行加密。

在本实施例中, 在得到改写后的问题语句之后, 可以对该文图语句进行意图识别, 判断该用户是否具有购买产品的意图; 若用户的意图为购买产品, 为了加快与用户之间的信息沟通, 则可以推荐对应所述指定意图的产品信息文本, 该产品信息文本中记载有所述用户意图购买产品的相关介绍信息, 且该产品信息文本中还具有输入用户信息的文本框。为了保障用户在上述产品信息文本中输入用户信息之后进行传输的安全性, 因此需要对其进行加密传输。在本实施例中, 约定对上述产品信息文本的加密密码为上述产品信息文本未输入用户信息之前的哈希值。

在一实施例中, 所述边权重的计算公式为:

$$weight(i, j) = P(w_{i+1}, w_{i+2}, \dots, w_j) = P(w_{i+1}|\widetilde{w_{i+1}})P(w_{i+2}|\widetilde{w_{i+2}}) \dots P(w_j|\widetilde{w_j})$$

其中, 所述问题语句中包括m个字, 所述问题语句表示为:

$$S = w_1, w_2, w_3, \dots, w_m$$

$P(w)$ 为问题语句中的字的概率分数, i, j 分别为两个节点对应所述问题语句中的位置; 事件 $\widetilde{w_j}$ 代表事件 $w_{j-n+1} \dots w_{j-1}$ 同时发生, n 为N-gram语言模型的窗口大小;

计算最短路径时, 为了使乘法变成加法, 此处边权重进行log处理, 因此, 上述最短路径的计算公式为:

$$P(w_1|\widetilde{w_1}) \dots P(w_m|\widetilde{w_m}) = \log(P(w_1|\widetilde{w_1})) + \log(P(w_m|\widetilde{w_m})).$$

在本实施例中, 以一个包括m个字的问题语句为例, 该问题语句表示为: $s = w_1, w_2, w_3, \dots, w_m$ 。

问题语句s的出现概率可表示为:

$$p(s) = P(w_1, w_2, w_3, \dots, w_m) = P(w_1)P(w_2|w_1) \dots P(w_m|w_1 \dots w_{m-1})$$

利用马尔科夫假设可以对上式进行近似求解, 以N-gram语言模型的窗口 $n=4$ 为例:

$$p(s) = P(w_1, w_2, w_3, \dots, w_m) = P(w_1)P(w_2|w_1) \dots P(w_m|w_{m-3} \dots w_{m-1})$$

将上式进行简化, 用事件 $\widetilde{w_m}$ 代表事件 $w_{m-3} \dots w_{m-1}$ 同时发生, 简化如下:

$$p(s) = P(w_1, w_2, w_3, \dots, w_m) = P(w_1)P(w_2|\widetilde{w_2}) \dots P(w_m|\widetilde{w_m})$$

定义有向图的节点集合为所有候选同义词集, 边集合为所有候选同义词与候选同义词间的n-gram拆解后的概率分数。

在一具体实施例中, 以N-gram语言模型的窗口 $n=4$ 为例, 上述问题语句为: 请问交费期间可以退保金鑫盛吗? 其中交费的候选同义词为“缴费”, 可以的候选同义词为“能不能”, 以计算上述“缴费”与“能不能”之间的边的边权重为例:

$$weight = P(w_5, w_6, \dots, w_9) = P(w_5|\widetilde{w}_5)P(w_6|\widetilde{w}_6) \dots P(w_9|\widetilde{w}_9)$$

上述问题语句的顶点（即起始节点、结束节点）至候选同义词的节点之间边的边权重计算以起始节点为例（距终止位的边权重计算方式同理），其计算公式如下：

$$weight = P(w_1, w_2, \dots, w_4) = P(w_1|\widetilde{w}_1)P(w_2|\widetilde{w}_2) \dots P(w_4|\widetilde{w}_4)$$

参照图2，本申请一实施例中还提供了一种问题语句的改写装置，包括：

获取单元10，用于获取用户输入的问题语句；

- 5 第一识别单元20，用于基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；

第一构建单元30，用于将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

- 10 第二构建单元40，用于将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

- 15 第一计算单元50，用于基于N-gram语言模型，计算所述有向图中各个边的边权重；

改写单元60，用于基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

在一实施例中，还包括：

- 20 第二计算单元，用于计算所述问题语句与所述改写后的问题语句的语义相似度；

判断单元，用于判断所述语义相似度是否大于阈值；

第三计算单元，用于若不大于，则基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出第二短路径，将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句；

- 25 匹配单元，用于若大于，则在数据库中匹配对应所述改写后的问题语句的答案，并将匹配出的所述答案推送至所述用户所在的用户终端。

在一实施例中，所述第二计算单元具体用于：

基于预设的编码器，得到所述问题语句对应的第一向量矩阵，以及得到所述改写后的问题语句对应的第二向量矩阵；

- 30 计算第一向量矩阵与所述第二向量矩阵的相似度。

相似度的计算公式为:

$$\text{Sim} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}};$$

其中, A_i 为所述第一向量矩阵的第 i 个分向量的数值, B_i 为所述第二向量矩阵的第 i 个分向量的数值, 所述第一向量矩阵和所述第二向量矩阵均包括 n 个分向量。

5 在一实施例中, 上述装置还包括:

第二识别单元, 用于基于所述改写后的问题语句, 识别所述用户的意图;

意图判断单元, 用于判断所述意图是否为指定意图; 所述指定意图为购买产品意图;

哈希计算单元, 用于若是, 则获取对应所述指定意图的产品信息文本, 并对所述产品信息文本进行哈希计算, 得到对应的哈希值存储于数据库中;

10 发送单元, 用于将所述产品信息文本以及所述哈希值发送至所述用户所在的用户终端, 并提醒所述用户在所述产品信息文本中输入用户信息, 并基于所述哈希值对所述产品信息文本进行加密。

在一实施例中, 所述边权重的计算公式为:

$$\text{weight}(i, j) = P(w_{i+1}, w_{i+2}, \dots, w_j) = P(w_{i+1} | \widetilde{w_{i+1}}) P(w_{i+2} | \widetilde{w_{i+2}}) \dots P(w_j | \widetilde{w_j})$$

其中, 所述问题语句中包括 m 个字, 所述问题语句表示为:

$$S = w_1, w_2, w_3, \dots, w_m$$

15 $P(w)$ 为问题语句中的字的概率分数, i, j 分别为两个节点对应应在所述问题语句中的位置; 事件 $\widetilde{w_j}$ 代表事件 $w_{j-n+1} \dots w_{j-1}$ 同时发生, n 为 N -gram 语言模型的窗口大小;

所述最短路径的计算公式为:

$$P(w_1 | \widetilde{w_1}) \dots P(w_m | \widetilde{w_m}) = \log(P(w_1 | \widetilde{w_1})) + \log(P(w_m | \widetilde{w_m})).$$

20 在本实施例中, 上述各个单元的具体实现请参照上述方法实施例中所述, 在此不再进行赘述。

参照图3, 本申请实施例中还提供一种计算机设备, 该计算机设备可以是服务器, 其内部结构可以如图3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。其中, 该计算机设计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、计算机程序
25 和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储问题语句等。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。该计算机程序被处理器执行时以实现一种问题语句的改写方法, 包括:

获取用户输入的问题语句；

基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；

5 将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

10 将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

基于N-gram语言模型，计算所述有向图中各个边的边权重；

15 基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

本领域技术人员可以理解，图3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。

本申请一实施例还提供一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现一种问题语句的改写方法，包括：

20 获取用户输入的问题语句；

基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；

将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

25 将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

30 节点；

基于N-gram语言模型，计算所述有向图中各个边的边权重；

基于所述有向图以及所述有向图中各个边的边权重，通过 N 最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

可以理解的是，本实施例中的计算机可读存储介质可以是易失性可读存储介质，也可以为 5 非易失性可读存储介质。

综上所述，为本申请实施例中提供的问题语句的改写方法、装置、计算机设备和存储介质，获取用户输入的问题语句；基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；将所述候选同义词之间的 10 句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；基于 N-gram 语言模型，计算所述有向图中各个边的边权重；通过 N 最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。本申请通过最短路径的计算，找出最符合语境同义词，从而对文图语句进行改写。

15 以上所述仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其它相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

1、一种问题语句的改写方法，其中，包括以下步骤：

获取用户输入的问题语句；

基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候

5 选同义词；

将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

10 将所述候选同义词之间的句子片段作为连接节点之间的边，分别在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

基于N-gram语言模型，计算所述有向图中各个边的边权重；

15 基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

2、根据权利要求1所述的问题语句的改写方法，其中，所述基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个

20 节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句的步骤之后，包括：

计算所述问题语句与所述改写后的问题语句的语义相似度；

判断所述语义相似度是否大于阈值；

25 若不大于，则基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出第二短路径，将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句；

若大于，则在数据库中匹配对应所述改写后的问题语句的答案，并将匹配出的所述答案推送至所述用户所在的用户终端。

3、根据权利要求2所述的问题语句的改写方法，其中，所述计算所述问题语句与所述改写后的问题语句的语义相似度的步骤，包括：

30 基于预设的编码器，得到所述问题语句对应的第一向量矩阵，以及得到所述改写后的问题语句对应的第二向量矩阵；

计算第一向量矩阵与所述第二向量矩阵的相似度。

相似度的计算公式为：

$$\text{Sim} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}};$$

其中， A_i 为所述第一向量矩阵的第*i*个分向量的数值， B_i 为所述第二向量矩阵的第*i*个分向量的数值，所述第一向量矩阵和所述第二向量矩阵均包括*n*个分向量。

4、根据权利要求1所述的问题语句的改写方法，其中，所述基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句的步骤之后，包括：

基于所述改写后的问题语句，识别所述用户的意图；

10 判断所述意图是否为指定意图；所述指定意图为购买产品意图；

若是，则获取对应所述指定意图的产品信息文本，并对所述产品信息文本进行哈希计算，得到对应的哈希值存储于数据库中；

将所述产品信息文本以及所述哈希值发送至所述用户所在的用户终端，并提醒所述用户在所述产品信息文本中输入用户信息，并基于所述哈希值对所述产品信息文本进行加密。

15 5、根据权利要求1所述的问题语句的改写方法，其中，所述边权重的计算公式为：

$$\text{weight}(i, j) = P(w_{i+1}, w_{i+2}, \dots, w_j) = P(w_{i+1} | \widetilde{w_{i+1}}) P(w_{i+2} | \widetilde{w_{i+2}}) \dots P(w_j | \widetilde{w_j})$$

其中，所述问题语句中包括*m*个字，所述问题语句表示为：

$$S = w_1, w_2, w_3, \dots, w_m$$

$P(w)$ 为问题语句中的字的概率分数，*i*，*j*分别为两个节点对应应在所述问题语句中的位置；事件 $\widetilde{w_j}$ 代表事件 $w_{j-n+1} \dots w_{j-1}$ 同时发生，*n*为N-gram语言模型的窗口大小；

所述最短路径的计算公式为：

$$20 \quad P(w_1 | \widetilde{w_1}) \dots P(w_m | \widetilde{w_m}) = \log(P(w_1 | \widetilde{w_1})) + \log(P(w_m | \widetilde{w_m})).$$

6、一种问题语句的改写装置，其中，包括：

获取单元，用于获取用户输入的问题语句；

第一识别单元，用于基于同义词匹配，识别所述问题语句中每个可被替换的词，以及替换每个所述词的候选同义词；

25 第一构建单元，用于将每个所述词的候选同义词作为节点，将所述问题语句的起点作为起始节点，所述问题语句的终点作为结束节点；

第二构建单元，用于将所述候选同义词之间的句子片段作为连接节点之间的边，分别

在所述起始节点、结束节点与所述节点之间构建边；基于所述起始节点、结束节点、各个所述节点以及所述边，构建有向图；其中，所述有向图中的各个节点对应的候选同义词之间不存在交叉，所述有向图中由起始节点开始，由结束节点结束，且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点；

第一计算单元，用于基于N-gram语言模型，计算所述有向图中各个边的边权重；

改写单元，用于基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句。

7、根据权利要求6所述的问题语句的改写装置，其中，还包括：

第二计算单元，用于计算所述问题语句与所述改写后的问题语句的语义相似度；

判断单元，用于判断所述语义相似度是否大于阈值；

第三计算单元，用于若不大于，则基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出第二短路径，将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句；

匹配单元，用于若大于，则在数据库中匹配对应所述改写后的问题语句的答案，并将匹配出的所述答案推送至所述用户所在的用户终端。

8、根据权利要求7所述的问题语句的改写装置，其中，所述第二计算单元具体用于：

基于预设的编码器，得到所述问题语句对应的第一向量矩阵，以及得到所述改写后的问题语句对应的第二向量矩阵；

计算第一向量矩阵与所述第二向量矩阵的相似度。

相似度的计算公式为：

$$\text{Sim} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}};$$

其中， A_i 为所述第一向量矩阵的第*i*个分向量的数值， B_i 为所述第二向量矩阵的第*i*个分向量的数值，所述第一向量矩阵和所述第二向量矩阵均包括*n*个分向量。

9、根据权利要求6所述的问题语句的改写装置，其中，还包括：第二识别单元，用于基于所述改写后的问题语句，识别所述用户的意图；

意图判断单元，用于判断所述意图是否为指定意图；所述指定意图为购买产品意图；

哈希计算单元，用于若是，则获取对应所述指定意图的产品信息文本，并对所述产品信息文本进行哈希计算，得到对应的哈希值存储于数据库中；

发送单元,用于将所述产品信息文本以及所述哈希值发送至所述用户所在的用户终端,并提醒所述用户在所述产品信息文本中输入用户信息,并基于所述哈希值对所述产品信息文本进行加密。

10、根据权利要求6所述的问题语句的改写装置,其中,所述边权重的计算公式为:

$$weight(i, j) = P(w_{i+1}, w_{i+2}, \dots, w_j) = P(w_{i+1}|\widetilde{w_{i+1}})P(w_{i+2}|\widetilde{w_{i+2}}) \dots P(w_j|\widetilde{w_j})$$

5 其中,所述问题语句中包括m个字,所述问题语句表示为:

$$S = w_1, w_2, w_3, \dots, w_m$$

$P(w)$ 为问题语句中的字的概率分数, i, j 分别为两个节点对应所述问题语句中的位置;事件 $\widetilde{w_j}$ 代表事件 $w_{j-n+1} \dots w_{j-1}$ 同时发生, n 为N-gram语言模型的窗口大小;

所述最短路径的计算公式为:

$$P(w_1|\widetilde{w_1}) \dots P(w_m|\widetilde{w_m}) = \log(P(w_1|\widetilde{w_1})) + \log(P(w_m|\widetilde{w_m})).$$

10 11、一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,其中,所述处理器执行所述计算机程序时实现一种问题语句的改写方法,包括:

获取用户输入的问题语句;

基于同义词匹配,识别所述问题语句中每个可被替换的词,以及替换每个所述词的候选同义词;

15 将每个所述词的候选同义词作为节点,将所述问题语句的起点作为起始节点,所述问题语句的终点作为结束节点;

20 将所述候选同义词之间的句子片段作为连接节点之间的边,分别在所述起始节点、结束节点与所述节点之间构建边;基于所述起始节点、结束节点、各个所述节点以及所述边,构建有向图;其中,所述有向图中的各个节点对应的候选同义词之间不存在交叉,所述有向图中由起始节点开始,由结束节点结束,且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点;

基于N-gram语言模型,计算所述有向图中各个边的边权重;

25 基于所述有向图以及所述有向图中各个边的边权重,通过N最短路径算法计算出最短路径,将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中,得到改写后的问题语句。

12、根据权利要求11所述的计算机设备,其中,所述基于所述有向图以及所述有向图中各个边的边权重,通过N最短路径算法计算出最短路径,将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中,得到改写后的问题语句的步骤之后,包括:

计算所述问题语句与所述改写后的问题语句的语义相似度;

判断所述语义相似度是否大于阈值;

若不大于,则基于所述有向图以及所述有向图中各个边的边权重,通过N最短路径算法计算出第二短路径,将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中,得到改写后的问题语句;

若大于,则在数据库中匹配对应所述改写后的问题语句的答案,并将匹配出的所述答案推送至所述用户所在的用户终端。

13、根据权利要求12所述的计算机设备,其中,所述计算所述问题语句与所述改写后的问题语句的语义相似度的步骤,包括:

10 基于预设的编码器,得到所述问题语句对应的第一向量矩阵,以及得到所述改写后的问题语句对应的第二向量矩阵;

计算第一向量矩阵与所述第二向量矩阵的相似度。

相似度的计算公式为:

$$\text{Sim} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}};$$

15 其中, A_i 为所述第一向量矩阵的第*i*个分向量的数值, B_i 为所述第二向量矩阵的第*i*个分向量的数值,所述第一向量矩阵和所述第二向量矩阵均包括*n*个分向量。

14、根据权利要求11所述的计算机设备,其中,所述基于所述有向图以及所述有向图中各个边的边权重,通过N最短路径算法计算出最短路径,将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中,得到改写后的问题语句的步骤之后,包括:

20 基于所述改写后的问题语句,识别所述用户的意图;

判断所述意图是否为指定意图;所述指定意图为购买产品意图;

若是,则获取对应所述指定意图的产品信息文本,并对所述产品信息文本进行哈希计算,得到对应的哈希值存储于数据库中;

25 将所述产品信息文本以及所述哈希值发送至所述用户所在的用户终端,并提醒所述用户在所述产品信息文本中输入用户信息,并基于所述哈希值对所述产品信息文本进行加密。

15、根据权利要求11所述的计算机设备,其中,所述边权重的计算公式为:

$$\text{weight}(i, j) = P(w_{i+1}, w_{i+2}, \dots, w_j) = P(w_{i+1} | \widetilde{w_{i+1}}) P(w_{i+2} | \widetilde{w_{i+2}}) \dots P(w_j | \widetilde{w_j})$$

其中,所述问题语句中包括*m*个字,所述问题语句表示为:

$$S = w_1, w_2, w_3, \dots, w_m$$

$P(w)$ 为问题语句中的字的概率分数, i, j 分别为两个节点对应在该问题语句中的位置; 事件 $\widetilde{w_j}$ 代表事件 $w_{j-n+1} \dots w_{j-1}$ 同时发生, n 为N-gram语言模型的窗口大小;

所述最短路径的计算公式为:

$$P(w_1|\widetilde{w_1}) \dots P(w_m|\widetilde{w_m}) = \log(P(w_1|\widetilde{w_1})) + \log(P(w_m|\widetilde{w_m})).$$

- 5 16、一种计算机可读存储介质, 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一种问题语句的改写方法, 包括:

获取用户输入的问题语句;

基于同义词匹配, 识别所述问题语句中每个可被替换的词, 以及替换每个所述词的候选同义词;

- 10 将每个所述词的候选同义词作为节点, 将所述问题语句的起点作为起始节点, 所述问题语句的终点作为结束节点;

将所述候选同义词之间的句子片段作为连接节点之间的边, 分别在所述起始节点、结束节点与所述节点之间构建边; 基于所述起始节点、结束节点、各个所述节点以及所述边, 构建有向图; 其中, 所述有向图中的各个节点对应的候选同义词之间不存在交叉, 所述有向图中由起始节点开始, 由结束节点结束, 且连接所述节点之间的边的方向为所述问题语句中排列在前的候选同义词对应的节点指向所述问题语句中排列在后的候选同义词对应的节点;

基于N-gram语言模型, 计算所述有向图中各个边的边权重;

- 20 基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句。

- 25 17、根据权利要求16所述的计算机可读存储介质, 其中, 所述基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出最短路径, 将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句的步骤之后, 包括:

计算所述问题语句与所述改写后的问题语句的语义相似度;

判断所述语义相似度是否大于阈值;

- 30 若不大于, 则基于所述有向图以及所述有向图中各个边的边权重, 通过N最短路径算法计算出第二短路径, 将所述第二短路径中各个节点对应的候选同义词改写至所述问题语句中, 得到改写后的问题语句;

若大于, 则在数据库中匹配对应所述改写后的问题语句的答案, 并将匹配出的所述答

案推送至所述用户所在的用户终端。

18、根据权利要求17所述的计算机可读存储介质，其中，所述计算所述问题语句与所述改写后的问题语句的语义相似度的步骤，包括：

5 基于预设的编码器，得到所述问题语句对应的第一向量矩阵，以及得到所述改写后的问题语句对应的第二向量矩阵；

计算第一向量矩阵与所述第二向量矩阵的相似度。

相似度的计算公式为：

$$\text{Sim} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}};$$

10 其中， A_i 为所述第一向量矩阵的第*i*个分向量的数值， B_i 为所述第二向量矩阵的第*i*个分向量的数值，所述第一向量矩阵和所述第二向量矩阵均包括*n*个分向量。

19、根据权利要求16所述的计算机可读存储介质，其中，所述基于所述有向图以及所述有向图中各个边的边权重，通过N最短路径算法计算出最短路径，将所述最短路径中各个节点对应的候选同义词改写至所述问题语句中，得到改写后的问题语句的步骤之后，包括：

15 基于所述改写后的问题语句，识别所述用户的意图；

判断所述意图是否为指定意图；所述指定意图为购买产品意图；

若是，则获取对应所述指定意图的产品信息文本，并对所述产品信息文本进行哈希计算，得到对应的哈希值存储于数据库中；

20 将所述产品信息文本以及所述哈希值发送至所述用户所在的用户终端，并提醒所述用户在所述产品信息文本中输入用户信息，并基于所述哈希值对所述产品信息文本进行加密。

20、根据权利要求16所述的计算机可读存储介质，其中，所述边权重的计算公式为：

$$\text{weight}(i, j) = P(w_{i+1}, w_{i+2}, \dots, w_j) = P(w_{i+1} | \widetilde{w_{i+1}}) P(w_{i+2} | \widetilde{w_{i+2}}) \dots P(w_j | \widetilde{w_j})$$

其中，所述问题语句中包括*m*个字，所述问题语句表示为：

$$S = w_1, w_2, w_3, \dots, w_m$$

$P(w)$ 为问题语句中的字的概率分数，*i*，*j*分别为两个节点对应所述问题语句中的位置；事件 $\widetilde{w_j}$ 代表事件 $w_{j-n+1} \dots w_{j-1}$ 同时发生，*n*为N-gram语言模型的窗口大小；

25 所述最短路径的计算公式为：

$$P(w_1 | \widetilde{w_1}) \dots P(w_m | \widetilde{w_m}) = \log(P(w_1 | \widetilde{w_1})) + \log(P(w_m | \widetilde{w_m})).$$

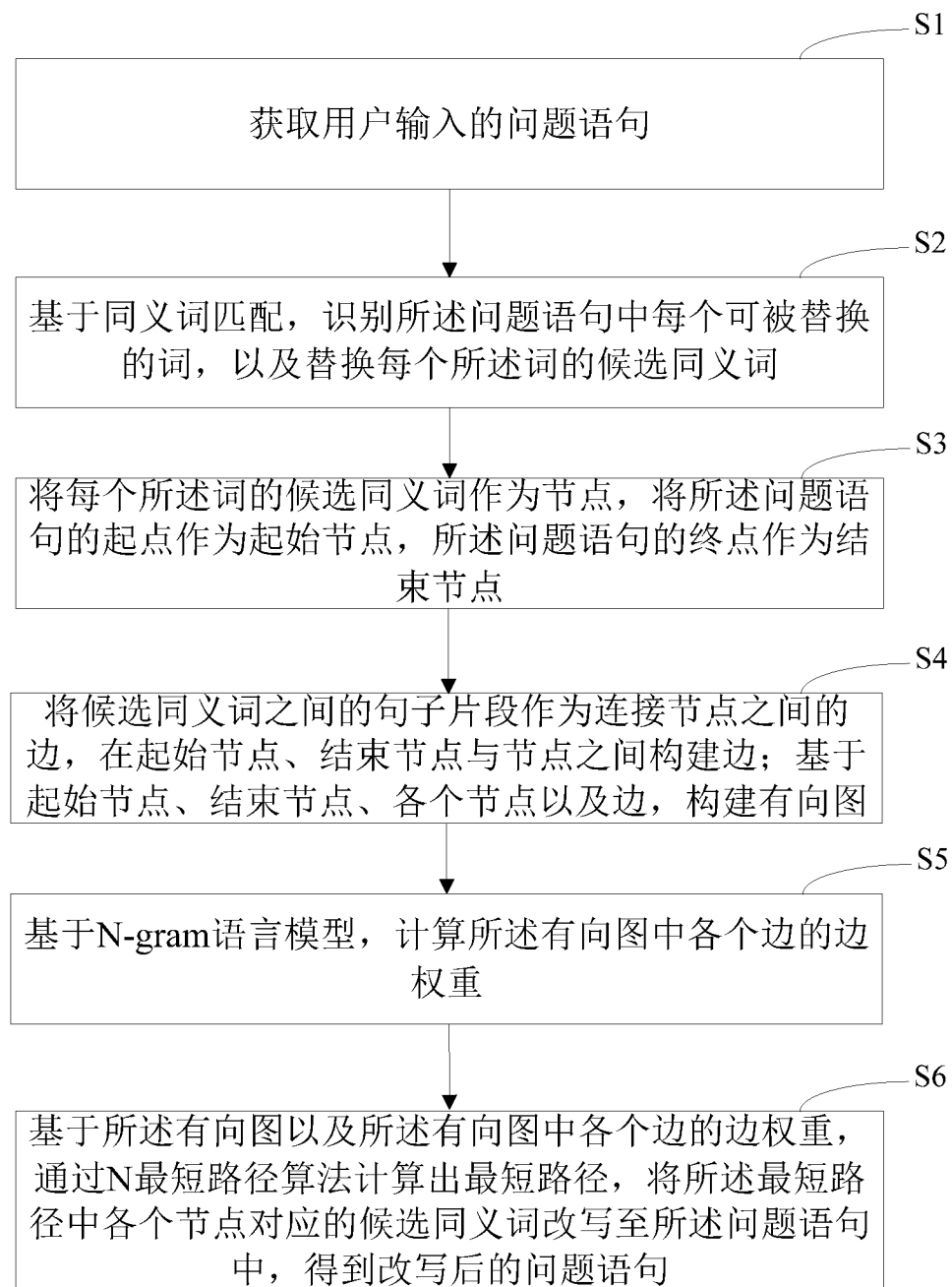


图 1



图 2

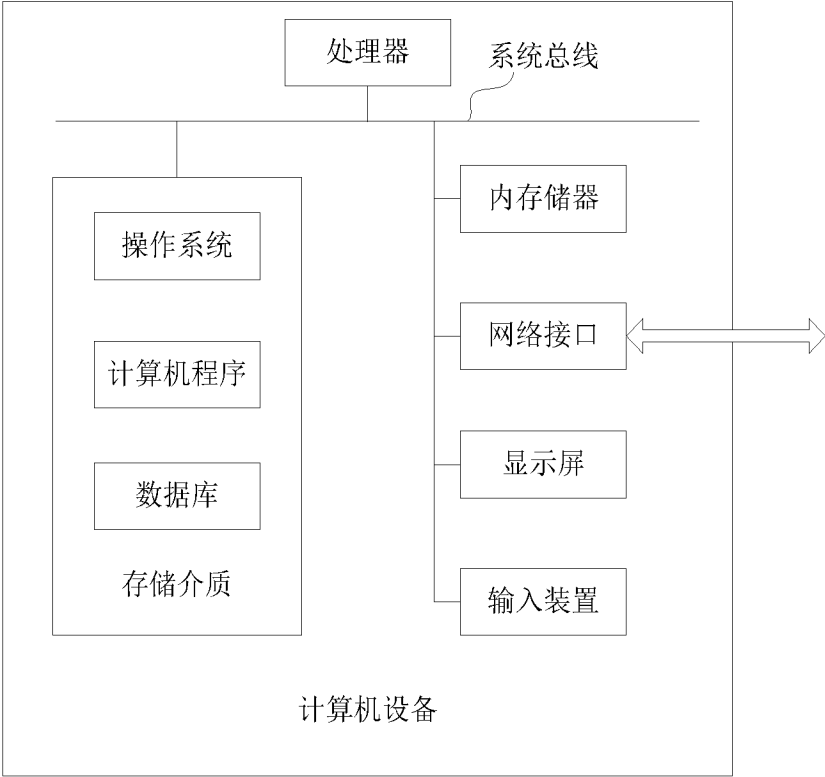


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/091700

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)

G06F16,G06F17/30

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)

DWPI; SIPOABS; CNABS; CNTXT; CNKI: 问题, 语句, 重写, 改写, 替换, 同义词, 近义词, 边, 图, 权重, 相似, 意图; question, query, sentence, rewrite, modify, replace, synonym, edge, graph, weight, similar, intension

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112287080 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 29 January 2021 (2021-01-29) description, paragraphs 8-58	1-20
Y	CN 110543555 A (ALIBABA GROUP HOLDING LIMITED) 06 December 2019 (2019-12-06) description, paragraphs 93-135, 141-142	1-20
Y	CN 105095178 A (ANHUI USTC IFLYTEK CO., LTD.) 25 November 2015 (2015-11-25) description paragraphs 74-114	1-20
A	WO 2020205047 A1 (MICROSOFT TECHNOLOGY LICENSING L.L.C.) 08 October 2020 (2020-10-08) 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

30 July 2021

Date of mailing of the international search report

05 August 2021

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/CN2021/091700

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112287080	A	29 January 2021	None			
CN	110543555	A	06 December 2019	None			
CN	105095178	A	25 November 2015	CN	105095178	B	14 August 2018
WO	2020205047	A1	08 October 2020	US	2020311070	A1	01 October 2020
				US	10970278	B2	06 April 2021

国际检索报告

国际申请号

PCT/CN2021/091700

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)

G06F16, G06F17/30

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))

DWPI; SIPOABS; CNABS; CNTXT; CNKI: 问题, 语句, 重写, 改写, 替换, 同义词, 近义词, 边, 图, 权重, 相似, 意图; question, query, sentence, rewrite, modify, replace, synonym, edge, graph, weight, similar, intension

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 112287080 A (平安科技深圳有限公司) 2021年 1月 29日 (2021 - 01 - 29) 说明书第8-58段	1-20
Y	CN 110543555 A (阿里巴巴集团控股有限公司) 2019年 12月 6日 (2019 - 12 - 06) 说明书第93-135, 141-142段	1-20
Y	CN 105095178 A (安徽科大讯飞信息科技股份有限公司) 2015年 11月 25日 (2015 - 11 - 25) 说明书第74-114段	1-20
A	WO 2020205047 A1 (MICROSOFT TECHNOLOGY LICENSING LLC) 2020年 10月 8日 (2020 - 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

2021年 7月 30日

国际检索报告邮寄日期

2021年 8月 5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

颜世莹

电话号码 62411767

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/091700

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	112287080	A	2021年 1月 29日	无			
CN	110543555	A	2019年 12月 6日	无			
CN	105095178	A	2015年 11月 25日	CN	105095178	B	2018年 8月 14日
WO	2020205047	A1	2020年 10月 8日	US	2020311070	A1	2020年 10月 1日
				US	10970278	B2	2021年 4月 6日