

# (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



(10) 国际公布号

(43) 国际公布日

2020 年 3 月 26 日 (26.03.2020)

WIPO | PCT

WO 2020/056979 A1

(51) 国际专利分类号:

G06F 16/955 (2019.01)

(21) 国际申请号:

PCT/CN2018/123597

(22) 国际申请日:

2018 年 12 月 25 日 (25.12.2018)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201811096202.5 2018年9月19日 (19.09.2018) 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: 周俊琨(ZHOU, JunKun); 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)。石卢靖(SHI, LuJing); 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)。许扬(XU, Yang); 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)。温利聪(WEN, LiCong); 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: 深圳市沃德知识产权代理有限公司(普通合伙)(SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市福田区园岭街道八卦四路 10 号中浩大厦 1528-1530 室于志光, 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,

(54) Title: KNOWLEDGE BASE SEARCH METHOD AND APPARATUS, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 知识库搜索方法、装置及计算机可读存储介质



(57) Abstract: The present application relates to big data technology. Disclosed is a knowledge base search method, comprising: acquiring a historical search record of a knowledge base matching a currently received keyword, and calculating historical operation behavior data of a knowledge point in the historical search record; calculating a weight coefficient of each knowledge point in the pre-set knowledge base; and according to the weight coefficient and the historical operation behavior data of the knowledge point, calculating, by means of a pre-set algorithm, the weight of each knowledge point found by using the currently received keyword, and according to the weights, sorting all the knowledge points found by using the currently received keyword and presenting same to a user. Further disclosed are a knowledge base search apparatus and a computer-readable storage medium. In the present application, the precise searching of a knowledge base is realized.

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))。

---

(57) 摘要: 本申请涉及一种大数据技术, 揭露了一种知识库搜索方法, 包括: 获取所述知识库与当前接收的关键字相匹配的历史搜索记录, 并计算所述历史搜索记录中的知识点的历史操作行为数据; 计算所述预设知识库中每一条知识点的权值系数; 根据所述权值系数以及对所述知识点的历史操作行为数据, 利用预设的算法, 计算利用所述当前接收的关键字搜索到的每一条知识点的权值, 并根据所述权值, 将利用所述当前接收的关键字搜索到所有知识点进行排序呈现给用户。本申请还提出一种知识库搜索装置以及一种计算机可读存储介质。本申请实现了知识库的精准搜索。

## 知识库搜索方法、装置及计算机可读存储介质

5

## 技术领域

同的知识库搜索方法、装置及计算机可读存储介质。

## 背景技术

15

传统的知识库搜索方案中，当接收到用户输入一个特定问题时，所采用的搜索算法大多是基于知识文本本身的相关性来进行搜索排序的，这种搜索排序结果可能并不能使用户快速、准确地找到问题的答案，给用户及/或客户带来不好的使用体验。

20

### 发明内容

目的在于当用户在知识库中进行搜索时，给用户呈现出精准的搜索结果。

30

获取所述知识库中与当前接收的关键字相匹配的历史搜索记录,并计算所述历史搜索记录中的知识点的历史操作行为数据:

计算所述知识库中每一条知识点的权值系数:

35

此外，为实现上述目的，本申请还提供一种知识库搜索装置，该装置包括存储器和处理器，所述存储器中存储有可在所述处理器上运行的知识库搜

索程序, 所述知识库搜索程序被所述处理器执行时实现如下步骤:

获取所述知识库中与当前接收的关键字相匹配的历史搜索记录, 并计算所述历史搜索记录中的知识点的历史操作行为数据;

计算所述知识库中每一条知识点的权值系数;

- 5 根据所述权值系数以及所述知识点的历史操作行为数据, 利用预设的算法, 计算利用所述当前接收的关键字搜索到的每一条知识点的权值, 并根据所述权值, 将利用所述当前接收的关键字搜索到所有知识点进行排序呈现给用户。

10 此外, 为实现上述目的, 本申请还提供一种计算机可读存储介质, 所述计算机可读存储介质上存储有知识库搜索程序, 所述知识库搜索程序可被一个或者多个处理器执行, 以实现如上所述的知识库搜索方法的步骤。

本申请提出的知识库搜索方法、装置及计算机可读存储介质, 在用户利用关键字在一个知识库中进行搜索时, 结合所述知识库中所有知识的内容权重以及所有用户在所述知识库中搜索时的操作行为, 对利用所述关键字搜索  
15 得到的知识点进行排序, 从而可以给用户呈现出精准的搜索结果。

## 附图说明

图1为本申请一实施例提供的知识库搜索方法的流程示意图;

图2为本申请一实施例提供的知识库搜索装置的内部结构示意图;

20 图3为本申请一实施例提供的知识库搜索装置中知识库搜索程序的模块示意图。

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。

## 25 具体实施方式

应当理解, 此处所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。

本申请提供一种知识库搜索方法。参照图 1 所示, 为本申请一实施例提供的知识库搜索方法的流程示意图。该方法可以由一个装置执行, 该装置可以  
30 可以由软件和/或硬件实现。

在本实施例中, 知识库搜索方法包括:

S1、接收用户输入到预设知识库的关键字, 并对接收的关键字执行分词操作。

本申请较佳实施例中, 所述用户可以是某一个金融企业, 如中国平安等的  
35 的客服人员。当所述客服人员通过电话、网页、APP、或者微信等渠道接收到一个用户询问时, 可以根据所述用户询问向该金融企业的知识库中输入一个关键字, 如“平安的尊宏人生有什么特点”。本申请较佳实施例在接收到所述关键字后, 对当前接收的关键字执行分词操作。

本申请较佳实施例采用长词优先原则, 根据预存的词库对接收的关键字

进行分词。

例如，所述长词优先原则指的是：对于一个需要分词的短语 T1，先从第一个字 A 开始，从预存的词库找出一个由 A 起始的最长词语 X1，然后从 T1 中剔除 X1 剩下 T2，再对 T2 采用相同的切分原理，切分后的结果为“X1/X2/.....”。

所述预存的词库可以包括常规的词库以及特定词库，如金融词库和产品词库等。例如，根据常规的词库，可以得到“平安”、“人生”、“什么”、“特点”等分词，但是根据金融词库，可以得到“尊宏人生”这个分词。因此，本申请较佳实施例对所述关键字“平安的尊宏人生有什么特点”进行分词后，可以得到的分词为“平安”、“的”、“尊宏人生”、“有”、“什么”、“特点”。

进一步地，本申请较佳实施例中，还可以进一步对得到的分词进行过滤处理，具体地，过滤处理采用以下任一种或两种方式：方式一：根据词性进行过滤，保留名词、动词以及形容词；方式二：根据频次进行过滤，保留频次大于频次阈值的分词，其中，频次是指分词在知识库中出现的频率或者次数。

S2、获取所有用户在所述预设知识库的历史搜索记录。

例如，用户 A1 以第一关键字进行搜索，得到数量为 N1 条的知识点；用户 A2 以第二关键字进行搜索，得到数量为 N2 条的知识点；用户 A3 以第三关键字进行搜索，得到数量为 N3 条的知识点等等，则获取包括每一次搜索的关键字以及对应得到的知识点的所有用户的历史搜索记录。

S3、将所述历史搜索记录的关键字执行分词操作，并与当前接收的关键字进行匹配，以查找到与当前接收的关键字匹配的历史搜索记录。

根据上述例子，本申请较佳实施例将第一关键字、第二关键字以及第三关键字按照上述介绍的方法进行分词操作。

进一步地，本申请较佳实施例将根据第一关键字得到的分词与上述根据当前接收的关键字得到的分词进行比较，得到两者的相似度，将根据第二关键字得到的分词与上述根据当前接收的关键字得到的分词进行比较，得到两者的相似度，以及将根据第三关键字得到的分词与上述根据当前接收的关键字得到的分词进行比较，得到两者的相似度。当相似度大于一个预设数值，如 50%，则认为该条历史搜索纪录当前接收的关键字匹配。

所述相似度的计算可以是比较两者相同的分词占所有分词的百分比。

在一个实施例中，根据当前接收的关键字，得到的分词为“平安”、“尊宏人生”、“什么”、“特点”，根据所述历史搜索记录中的第一关键字，得到的分词为“平安”、“尊宏人生”、“特点”，则两者的相似度为75%。

进一步地，在本申请其他实施例中，也可以根据一些特定的词库，如金融词库和产品词库，赋予每个分词不同的权值。例如，“尊宏人生”为来于自产品词库，则可以赋予该分词 1.5 的权值等。

S4、获取相匹配的历史搜索记录被执行的预设操作行为。

本申请较佳实施例中，所述预设操作行为为对所述历史搜索记录中的知

识点的点击行为。

在一个实施例中，根据上述匹配，用户 A1 以第一关键字进行搜索，得到数量为 N1 条的知识点，以及以第二关键字进行搜索，得到数量为 N2 条的知识点为相匹配的历史搜索记录，则获取该条历史搜索记录的点击行为。

5        假设，以第一关键字进行搜索，得到的 N1 条的知识点包括 t1、t2、t3、t4、t5、t6、t7、t8、t9 以及 t10，其中用户通过点击行为依次查看了知识点 t1、t3、t4 以及 t8，则获取的点击行为依次包括所述知识点 t1、t3、t4 以及 t8；以第二关键字进行搜索，得到的 N2 条的知识点包括 t1、t11、t2、t5、t12、t13、t8、t14，以及 t15，其中用户通过点击行为依次查看了知识点 t1、t11、t5 以及  
10        t8，则获取的点击行为依次包括所述知识点 t1、t11、t5 以及 t8。

S5、统计每个相匹配的历史搜索记录中最后一条被执行预设操作行为的知识点，并计算所述每个最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比。

根据上述例子，以第一关键字进行搜索，得到的 N1 条的知识点包括 t1、  
15        t2、t3、t4、t5、t6、t7、t8、t9 以及 t10，其中用户通过点击行为依次查看了知识点 t1、t3、t4 以及 t8，则最后一条被点击的知识点为 t8。

本案其中一个实施例根据上述流程，共得到三条相匹配的历史搜索记录，其中，第一条相匹配的历史搜索记录中，t8 为用户最后一次点击的知识点，第二条相匹配的历史搜索记录中，t15 为用户最后一次点击的知识点，以及第  
20        三条相匹配的历史搜索记录中，t8 为用户最后一次点击的知识点，则知识点 t8 的百分比为 67%，知识点 t15 的百分比为 33%。

应该了解，用户最后一次点击的知识点可能代表该知识点解决了用户的问题。

S6、统计每个相匹配的历史搜索记录中除去最后一条知识点的其他条被执行预设操作行为的知识点，并计算其中每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比。

根据上述例子，以第一关键字进行搜索，得到的 N1 条的知识点包括 t1、t2、t3、t4、t5、t6、t7、t8、t9 以及 t10，其中用户通过点击行为依次查看了  
30        知识点 t1、t3、t4 以及 t8，则除去最后一条知识点的其他条被点击的知识点依次为 t1、t3、t4。

本申请其中一个实施例根据上述流程，共得到三条相匹配的历史搜索记录，其中，第一条相匹配的历史搜索记录中，除去最后一条知识点的其他条被点击的知识点为 t1、t3、t4；第二条相匹配的历史搜索记录中，除去最后一条知识点的其他条被点击的知识点为 t1、t5、t6；以及第三条相匹配的历史搜索记录中，除去最后一条知识点的其他条被点击的知识点为 t3、t7、t8；，则  
35        知识点 t1 的第二百分比为 67%、知识点 t3 的第二百分比为 67%、知识点 t4、t5、t6、t7、t8 的第二百分比均为 33%。

S7、获取所述预设知识库中的所有知识点，并对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分

词。

优选地，所述预先确定的分词规则如下：

按预设类型标点符号（例如，“、”、“。”、“！”、“；”等等）对获取的各个信息进行短句拆分（例如，从信息起始位置至第一个预设类型标点符号之间的信息为一个短句；若信息结束位置无预设类型标点符号，则从倒数第一个预设类型标点符号至信息结束位置之间的信息为一个短句，且针对从第一个预设类型标点符号至倒数第一个预设类型标点符号之间的信息，每两个预设类型标点符号之间的信息为一个短句；若信息结束位置有预设类型标点符号，则针对从第一个预设类型标点符号至倒数第一个预设类型标点符号之间的信息，每两个预设类型标点符号之间的信息为一个短句）；

对拆分的每一个短句，采用长词优先原则，根据预设的词库进行分词（例如，长词优先原则指的是：对于一个需要分词的短语 T1，先从第一个字 A 开始，从预存的词库找出一个由 A 起始的最长词语 X1，然后从 T1 中剔除 X1 剩下 T2，再对 T2 采用相同的切分原理，切分后的结果为“X1/X2/\、\、\、\、\、\、\、”；例如，在预存的词库中包括“平安”、“推出”、“了”、“尊宏人生”和“产品”时，短语“平安推出了尊宏人生产品”的切分结果为“平安”/“推出”/“了”/“尊宏人生”/“产品”）。

进一步地，本申请较佳实施例，还可以进一步对得到的分词进行过滤处理，具体地，过滤处理采用以下任一种或两种方式：方式一：根据词性进行过滤，保留名词、动词以及形容词；方式二：根据频次进行过滤，保留频次大于频次阈值的分词，其中，频次是指分词在知识库中出现的频率或者次数。

本申请较佳实施例中，所述预存的词库包括常规的词库，以及特定词库，如金融词库和产品词库等。

S8、计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

本申请较佳实施例中，计算所述每个知识点中的文档分词的词频（Tf）、反词频（Idf）、词权重（Tw）以及位置权重（Fw），并计算所述每一个知识点的知识权重（Kw）、长度归一化因子（Norm）等。

其中：

所述词频（Tf）指的是每个文档分词在包含该文档分词的知识点中出现的次数；

所述反词频（Idf）指的是每个文档分词在整个资料库中出现的次数；

词权重（Tw）是指每个文档分词的来源不同，其词权重不同，例如，若某一个文档分词来自于特定词库，如金融词库和产品词库等，则该文档分词的权重较高，例如设置为 1.2，非来自于所述特定词库的文档分词的权重较低，例如设置为 1；

位置权重（Fw）是指每个文档分词在知识点中的位置不同，则其权重不同，例如，若某一个文档分词位于知识点的标题中，则所述文档分词的位置权重较高，例如设置为 1.2，若某一个文档分词位于知识点的正文中，则所述

文档分词的位置权重较低,例如设置为 1,以及若某一个文档分词位于知识点的附件中,则所述文档分词的位置权重更低,例如设置为 0.8;

所述知识点的知识权重 (Kw) 为所述知识点中每个文档分词的词频-反词频的加权求和,即  $Kw = Tf1 * Idf1 + Tf2 * Idf2 + \dots + Tfn * Idfn$ ;

5 所述知识点的长度归一化因子  $Norm = 1 / (1 + \text{知识点的文本长度} / 1000)$ ;

S9, 根据上述每一条知识点以及其中的文档分词的一种或者多种预设的权值系数,以及所述每个最后一条被执行预设操作行为的知识点占有所有最后一条被执行预设操作行为的知识点的百分比及每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比,利用预设的算法,计算  
10 每一条知识点的权值 Weight,并根据所述权值,将所有知识点进行排序呈现给用户。

本申请较佳实施例中,所述预设的算法为:

$$Weight = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1+Idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw$$

其中:

15 Nd 为知识点的个数, i 为第几个分词, n 为共 n 个分词;

Lbw 为最后一次操作行为权重,是根据最后一条被点击的知识点的百分比设置的。

例如,一条知识点被最后一条被点击的百分比大于 80%,可以设置该知识点的权重为 1.4;一条知识点被最后一条被点击的百分比大于 50%但是小于  
20 等于 80%,可以设置该知识点的权重为 1.3;一条知识点被最后一条被点击的百分比大于 30%但是小于等于 50%,可以设置该知识点的权重为 1.2;一条知识点被最后一条被点击的百分比小于 30%,可以设置该知识点的权重为 1.1,否则,若一条知识点重来没有被最有一条点击过,则其权重为 1;

Bw 为操作行为权重,是根据被点击但非最后一条被点击的知识点的百分比  
25 设置的。

例如,一条知识点被被点击的百分比大于 80%,可以设置该知识点的权重为 1.2;一条知识点被被点击的百分比大于 50%但是小于等于 80%,可以设置  
30 该知识点的权重为 1.15;一条知识点被点击的百分比大于 30%但是小于等于 50%,可以设置该知识点的权重为 1.1;一条知识点被点击的百分比小于 30%,可以设置该知识点的权重为 1.05,否则,若一条知识点重来没有被点击过,则其权重为 1。

本申请其他实施例中,所述预设的算法也可以为:

$$Weight = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1+Idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw * Cw * W$$

其中:

35 Cw 是每一条知识点的渠道权重;

例如,当某一个知识点与当前搜索属于同一渠道,如同属客服人工搜索、或同属机器人智能搜索等,则所述知识点的渠道权重(Cw)设置为 1.2,否则,若不属于相同的渠道,则设置为 1.0。

W 是每一条知识点的所属公司权重；

例如，当某一个知识点与当前搜索属于同一公司，如同属寿险组公司，所述知识点的公司权重（W）设置为 1.2，否则，若不属于相同的公司，则设置为 1.0。

5

发明还提供一种知识库搜索装置。参照图 2 所示，为本申请一实施例提供的知识库搜索装置的内部结构示意图。

在本实施例中，所述知识库搜索装置 1 可以是 PC（Personal Computer，个人电脑），或者是智能手机、平板电脑、便携计算机等终端设备，也可以是一种服务器等。该知识库搜索装置 1 至少包括存储器 11、处理器 12，通信总线 13，以及网络接口 14。

其中，存储器 11 至少包括一种类型的可读存储介质，所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器（例如，SD 或 DX 存储器等）、磁性存储器、磁盘、光盘等。存储器 11 在一些实施例中可以是知识库搜索装置 1 的内部存储单元，例如该知识库搜索装置 1 的硬盘。存储器 11 在另一些实施例中也可以是知识库搜索装置 1 的外部存储设备，例如知识库搜索装置 1 上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。进一步地，存储器 11 还可以既包括知识库搜索装置 1 的内部存储单元也包括外部存储设备。存储器 11 不仅可以用于存储安装于知识库搜索装置 1 的应用软件及各类数据，例如知识库搜索程序 01 的代码等，还可以用于暂时地存储已经输出或者将要输出的数据。

处理器 12 在一些实施例中可以是一中央处理器（Central Processing Unit, CPU）、控制器、微控制器、微处理器或其他数据处理芯片，用于运行存储器 11 中存储的程序代码或处理数据，例如执行知识库搜索程序 01 等。

通信总线 13 用于实现这些组件之间的连接通信。

网络接口 14 可选的可以包括标准的有线接口、无线接口（如 WI-FI 接口），通常用于在该装置 1 与其他电子设备之间建立通信连接。

可选地，该装置 1 还可以包括用户接口，用户接口可以包括显示器（Display）、输入单元比如键盘（Keyboard），可选的用户接口还可以包括标准的有线接口、无线接口。可选地，在一些实施例中，显示器可以是 LED 显示器、液晶显示器、触控式液晶显示器以及 OLED（Organic Light-Emitting Diode，有机发光二极管）触摸器等。其中，显示器也可以适当的称为显示屏或显示单元，用于显示在知识库搜索装置 1 中处理的信息以及用于显示可视化的用户界面。

图 2 仅示出了具有组件 11-14 以及知识库搜索程序 01 的知识库搜索装置 1，本领域技术人员可以理解的是，图 1 示出的结构并不构成对知识库搜索装置 1 的限定，可以包括比图示更少或者更多的部件，或者组合某些部件，或者不同的部件布置。

在图 2 所示的装置 1 实施例中，存储器 11 中存储有知识库搜索程序 01；

处理器 12 执行存储器 11 中存储的知识库搜索程序 01 时实现如下步骤:

步骤一、接收用户输入到预设知识库的关键字，并对接收的关键字执行分词操作。

本申请较佳实施例中，所述用户可以是某一个金融企业，如中国平安等的客服人员。当所述客服人员通过电话、网页、APP、或者微信等渠道接收到一个用户询问时，可以根据所述用户询问向该金融企业的知识库中输入一个关键字，如“平安的尊宏人生有什么特点”。本申请较佳实施例在接收到所述关键字后，对当前接收的关键字执行分词操作。

本申请较佳实施例，所述分词操作采用长词优先原则，根据预存的词库  
10 对接收的关键字进行分词。

例如，所述长词优先原则指的是：对于一个需要分词的短语 T1，先从第一个字 A 开始，从预存的词库找出一个由 A 起始的最长词语 X1，然后从 T1 中剔除 X1 剩下 T2，再对 T2 采用相同的切分原理，切分后的结果为“X1/X2/、、、、、”。

15 所述预存的词库可以包括常规的词库以及特定词库，如金融词库和产品词库等。例如，根据常规的词库，可以得到“平安”、“人生”、“什么”、“特点”等分词，但是根据金融词库，可以得到“尊宏人生”这个分词。因此，本申请较佳实施例对所述关键字“平安的尊宏人生有什么特点”进行分词后，可以得到的分词为“平安”、“的”、“尊宏人生”、“有”、“什么”、“特点”。

进一步地，本申请较佳实施例中，还可以进一步对得到的分词进行过滤处理，具体地，过滤处理采用以下任一种或两种方式：方式一：根据词性进行过滤，保留名词、动词以及形容词；方式二：根据频次进行过滤，保留频次大于频次阈值的分词，其中，频次是指分词在知识库中出现的频率或者次数。

25 步骤二、获取所有用户在所述预设知识库的历史搜索记录。

例如，用户 A1 以第一关键字进行搜索，得到数量为 N1 条的知识点；用户 A2 以第二关键字进行搜索，得到数量为 N2 条的知识点；用户 A3 以第三关键字进行搜索，得到数量为 N3 条的知识点等等，则获取包括每一次搜索的关键字以及对应得到的知识点的所有用户的历史搜索记录。

30 步骤三、将所述历史搜索记录的关键字执行分词操作，并与当前接收的  
关键字进行匹配，以查找到与当前接收的关键字匹配的历史搜索记录。

根据上述例子，本申请较佳实施例将第一关键字、第二关键字以及第三关键字按照上述介绍的方法进行分词操作。

35 进一步地，本申请较佳实施例将根据第一关键字得到的分词与上述根据当前接收的关键字得到的分词进行比较，得到两者的相似度，将根据第二关键字得到的分词与上述根据当前接收的关键字得到的分词进行比较，得到两者的相似度，以及将根据第三关键字得到的分词与上述根据当前接收的关键字得到的分词进行比较，得到两者的相似度。当相似度大于一个预设数值，如 50%，则认为该条历史搜索纪录当前接收的关键字匹配。

所述相似度的计算可以是比较两者相同的分词占有所有分词的百分比。

在一个实施例中，根据当前接收的关键字，得到的分词为“平安”、“尊宏人生”、“什么”、“特点”，根据所述历史搜索记录中的第一关键字，得到的分词为“平安”、“尊宏人生”、“特点”，则两者的相似度为 75%。

5 进一步地，在本申请其他实施例中，也可以根据一些特定的词库，如金融词库和产品词库，赋予每个分词不同的权值。例如，“尊宏人生”为来于自产品词库，则可以赋予该分词 1.5 的权值等。

步骤四、获取相匹配的历史搜索记录被执行的预设操作行为。

10 本申请较佳实施例中，所述预设操作行为为对所述历史搜索记录中的知识点的点击行为。

在一个实施例中，根据上述匹配，用户 A1 以第一关键字进行搜索，得到数量为 N1 条的知识点，以及以第二关键字进行搜索，得到数量为 N2 条的知识点为相匹配的历史搜索记录，则获取该条历史搜索记录的点击行为。

15 假设，以第一关键字进行搜索，得到的 N1 条的知识点包括 t1、t2、t3、t4、t5、t6、t7、t8、t9 以及 t10，其中用户通过点击行为依次查看了知识点 t1、t3、t4 以及 t8，则获取的点击行为依次包括所述知识点 t1、t3、t4 以及 t8；以第二关键字进行搜索，得到的 N2 条的知识点包括 t1、t11、t2、t5、t12、t13、t8、t14，以及 t15，其中用户通过点击行为依次查看了知识点 t1、t11、t5 以及 t8，则获取的点击行为依次包括所述知识点 t1、t11、t5 以及 t8。

20 步骤五、统计每个相匹配的历史搜索记录中最后一条被执行预设操作行为的知识点，并计算所述每个最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比。

25 根据上述例子，以第一关键字进行搜索，得到的 N1 条的知识点包括 t1、t2、t3、t4、t5、t6、t7、t8、t9 以及 t10，其中用户通过点击行为依次查看了知识点 t1、t3、t4 以及 t8，则最后一条被点击的知识点为 t8。

30 本案其中一个实施例根据上述流程，共得到三条相匹配的历史搜索记录，其中，第一条相匹配的历史搜索记录中，t8 为用户最后一次点击的知识点，第二条相匹配的历史搜索记录中，t15 为用户最后一次点击的知识点，以及第三条相匹配的历史搜索记录中，t8 为用户最后一次点击的知识点，则知识点 t8 的百分比为 67%，知识点 t15 的百分比为 33%。

应该了解，用户最后一次点击的知识点可能代表该知识点解决了用户的问题。

35 步骤六、统计每个相匹配的历史搜索记录中除去最后一条知识点的其他条被执行预设操作行为的知识点，并计算其中每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比。

根据上述例子，以第一关键字进行搜索，得到的 N1 条的知识点包括 t1、t2、t3、t4、t5、t6、t7、t8、t9 以及 t10，其中用户通过点击行为依次查看了知识点 t1、t3、t4 以及 t8，则除去最后一条知识点的其他条被点击的知识点依次为 t1、t3、t4。

本申请其中一个实施例根据上述流程，共得到三条相匹配的历史搜索记录，其中，第一条相匹配的历史搜索记录中，除去最后一条知识点的其他条被点击的知识点为 t1、t3、t4；第二条相匹配的历史搜索记录中，除去最后一条知识点的其他条被点击的知识点为 t1、t5、t6；以及第三条相匹配的历史搜索记录中，除去最后一条知识点的其他条被点击的知识点为 t3、t7、t8；，则知识点 t1 的第二百分比为 67%、知识点 t3 的第二百分比为 67%、知识点 t4、t5、t6、t7、t8 的第二百分比均为 33%。

步骤七、获取所述预设知识库中的所有知识点，并对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词。

优选地，所述预先确定的分词规则如下：

按预设类型标点符号（例如，“、”、“。”、“！”、“；”等等）对知识点中的内容进行短句拆分（例如，从信息起始位置至第一个预设类型标点符号之间的信息为一个短句；若信息结束位置无预设类型标点符号，则从倒数第一个预设类型标点符号至信息结束位置之间的信息为一个短句，且针对从第一个预设类型标点符号至倒数第一个预设类型标点符号之间的信息，每两个预设类型标点符号之间的信息为一个短句；若信息结束位置有预设类型标点符号，则针对从第一个预设类型标点符号至倒数第一个预设类型标点符号之间的信息，每两个预设类型标点符号之间的信息为一个短句）；

对拆分的每一个短句，采用长词优先原则，根据预设的词库进行分词（例如，长词优先原则指的是：对于一个需要分词的短语 T1，先从第一个字 A 开始，从预存的词库找出一个由 A 起始的最长词语 X1，然后从 T1 中剔除 X1 剩下 T2，再对 T2 采用相同的切分原理，切分后的结果为“X1/X2/、……、”；例如，在预存的词库中包括“平安”、“推出”、“了”、“尊宏人生”和“产品”时，短语“平安推出了尊宏人生产品”的切分结果为“平安”/“推出”/“了”/“尊宏人生”/“产品”）。

进一步地，本申请较佳实施例中，还可以进一步对得到的分词进行过滤处理，具体地，过滤处理采用以下任一种或两种方式：方式一：根据词性进行过滤，保留名词、动词以及形容词；方式二：根据频次进行过滤，保留频次大于频次阈值的分词，其中，频次是指分词在知识库中出现的频率或者次数。

本申请较佳实施例中，所述预存的词库包括常规的词库，以及特定词库，如金融词库和产品词库等。

步骤八、计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

本申请较佳实施例中，计算所述每个知识点中的文档分词的词频 Tf、反词频 Idf、词权重 Tw 以及位置权重 Fw，并计算所述每一个知识点的知识权重 Kw、长度归一化因子 Norm 等。

其中：

所述词频 Tf 指的是每个文档分词在包含该文档分词的知识点中出现的次数;

所述反词频 Idf 指的是每个文档分词在整个资料库中出现的次数;

词权重 Tw 是指每个文档分词的来源不同, 其词权重不同, 例如, 若某一个文档分词来自于特定词库, 如金融词库和产品词库等, 则该文档分词的权重较高, 例如设置为 1.2, 非来自于所述特定词库的文档分词的权重较低, 例如设置为 1;

位置权重 Fw 是指每个文档分词在知识点中的位置不同, 则其权重不同, 例如, 若某一个文档分词位于知识点的标题中, 则所述文档分词的位置权重较高, 例如设置为 1.2, 若某一个文档分词位于知识点的正文中, 则所述文档分词的位置权重较低, 例如设置为 1, 以及若某一个文档分词位于知识点的附件中, 则所述文档分词的位置权重更低, 例如设置为 0.8;

所述知识点的知识权重 Kw 为所述知识点中每个文档分词的词频-反词频的加权求和, 即  $Kw = Tf_1 * Idf_1 + Tf_2 * Idf_2 + \dots + Tf_n * Idf_n$ ;

所述知识点的长度归一化因子  $Norm = 1 / (1 + \text{知识点的文本长度} / 1000)$ ;

步骤九、根据上述每一条知识点以及其中的文档分词的一种或者多种预设的权值系数, 以及所述每个最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比及每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比, 利用预设的算法, 计算每一条知识点的权值 Weight, 并根据所述权值, 将所有知识点进行排序呈现给用户。

本申请较佳实施例中, 所述预设的算法为:

$$Weight = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1 + Idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw。$$

其中:

Nd 为知识点的个数, i 为第几个分词, n 为共 n 个分词;

Lbw 为最后一次操作行为权重, 是根据最后一条被点击的知识点的百分比设置的。

例如, 一条知识点被最后一条被点击的百分比大于 80%, 可以设置该知识点的权重为 1.4; 一条知识点被最后一条被点击的百分比大于 50%但是小于等于 80%, 可以设置该知识点的权重为 1.3; 一条知识点被最后一条被点击的百分比大于 30%但是小于等于 50%, 可以设置该知识点的权重为 1.2; 一条知识点被最后一条被点击的百分比小于 30%, 可以设置该知识点的权重为 1.1, 否则, 若一条知识点重来没有被最有一条点击过, 则其权重为 1;

Bw 为操作行为权重, 是根据被点击但非最后一条被点击的知识点的百分比设置的。

例如, 一条知识点被被点击的百分比大于 80%, 可以设置该知识点的权重为 1.2; 一条知识点被被点击的百分比大于 50%但是小于等于 80%, 可以设置该知识点的权重为 1.15; 一条知识点被点击的百分比大于 30%但是小于等

于 50%，可以设置该知识点的权重为 1.1；一条知识点被点击的百分比小于 30%，可以设置该知识点的权重为 1.05，否则，若一条知识点重来没有被点击过，则其权重为 1。

本申请其他实施例中，所述预设的算法也可以为：

$$5 \quad \text{Weight} = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1+idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw * Cw * W$$

其中：

Cw 是每一条知识点的渠道权重；

例如，当某一个知识点与当前搜索属于同一渠道，如同属客服人工搜索、或同属机器人智能搜索等，则所述知识点的渠道权重(Cw)设置为 1.2，否则，  
10 若不属于相同的渠道，则设置为 1.0。

W 是每一条知识点的所属公司权重；

例如，当某一个知识点与当前搜索属于同一公司，如同属寿险组公司，所述知识点的公司权重(W)设置为 1.2，否则，若不属于相同的公司，则设置为 1.0。

15 可选地，在其他实施例中，知识库搜索程序还可以被分割为一个或者多个模块，一个或者多个模块被存储于存储器 11 中，并由一个或多个处理器(本实施例为处理器 12)所执行以完成本申请，本申请所称的模块是指能够完成特定功能的一系列计算机程序指令段，用于描述知识库搜索程序在知识库搜索装置中的执行过程。

例如，参照图 3 所示，为本申请知识库搜索装置一实施例中的知识库搜索程序的程序模块示意图，该实施例中，所述知识库搜索程序可以被分割为关键字接收模块 10、用户行为分析模块 20、知识点内容分析模块 30 以及搜索模块 40，示例性地：

25 所述关键字接收模块 10 用于：接收用户输入到预设知识库的关键字，并对接收的关键字执行分词操作。

所述用户行为分析模块 20 用于：获取所有用户在所述预设知识库的历史搜索记录，将所述历史搜索记录的关键字执行分词操作，并与当前接收的关键字进行匹配，以查找到与当前接收的关键字匹配的历史搜索记录，并计算  
30 所述历史搜索记录中的知识点的历史操作行为数据。

所述知识点内容分析模块 30 用于：获取所述预设知识库中的所有知识点，并对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词，并计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

35 所述搜索模块 40 用于：根据所述权值系数以及对所述知识点的历史操作行为数据，利用预设的算法，计算利用所述关键字搜索到的每一条知识点的权值，并根据所述权值，将利用所述关键字搜索到所有知识点进行排序呈现给用户。

上述关键字接收模块 10、用户行为分析模块 20、知识点内容分析模块 30

以及搜索模块 40 等程序模块被执行时所实现的功能或操作步骤与上述实施例大体相同，在此不再赘述。

此外，本申请实施例还提出一种计算机可读存储介质，所述计算机可读存储介质上存储有知识库搜索程序，所述知识库搜索程序可被一个或多个处理器执行，以实现如下操作：

接收用户输入到预设知识库的关键字，并对接收的关键字执行分词操作；

获取所有用户在所述预设知识库的历史搜索记录，将所述历史搜索记录的关键字执行分词操作，并与当前接收的关键字进行匹配，以查找到与当前接收的关键字匹配的历史搜索记录，并计算所述历史搜索记录中的知识点的历史操作行为数据；

获取所述预设知识库中的所有知识点，并对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词，并计算每一条知识点以及其中的文档分词的一种或者多种权值系数；

根据所述权值系数以及对所述知识点的历史操作行为数据，利用预设的算法，计算利用所述关键字搜索到的每一条知识点的权值，并根据所述权值，将利用所述关键字搜索到所有知识点进行排序呈现给用户。

本申请计算机可读存储介质具体实施方式与上述知识库搜索装置和方法各实施例基本相同，在此不作累述。

需要说明的是，上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。并且本文中的术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

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上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在如上所述的一个存储介质(如 ROM/RAM、磁碟、光盘)中，包括若干指令用以使得一台终端设备(可以是手机，计算机，服务器，或者网络设备等)执行本申请各个实施例所述的方法。

以上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他相关的技术领域，均同理包括在本申请的专利保护范围内。

# 权 利 要 求 书

1、一种知识库搜索方法，其特征在于，所述方法包括：

获取所述知识库中与当前接收的关键字相匹配的历史搜索记录，并计算  
5 所述历史搜索记录中的知识点的历史操作行为数据；

计算所述知识库中每一条知识点的权值系数；

根据所述权值系数以及所述知识点的历史操作行为数据，利用预设的算  
法，计算利用所述当前接收的关键字搜索到的每一条知识点的权值，并根据  
所述权值，将利用所述当前接收的关键字搜索到所有知识点进行排序呈现给  
10 用户。

2、如权利要求 1 所述的知识库搜索方法，其特征在于，所述获取所述知  
识库中与当前接收的关键字相匹配的历史搜索记录，并计算所述历史搜索记  
录中的知识点的历史操作行为数据，包括：

接收用户输入到所述知识库的关键字，并对接收的关键字执行分词操作；

15 获取所有用户在所述知识库的历史搜索记录，将所述历史搜索记录的关  
键字执行分词操作，并与当前接收的关键字进行匹配，以查找到与当前接收  
的关键字匹配的历史搜索记录；及

计算所述历史搜索记录中的知识点的历史操作行为数据。

3、如权利要求 2 所述的知识库搜索方法，其特征在于，所述计算所述历  
20 史搜索记录中的知识点的历史操作行为数据包括：

根据相匹配的历史搜索记录被执行的预设操作行为数据，统计每个相匹  
配的历史搜索记录中最后一条被执行预设操作行为的知识点；

计算所述每个最后一条被执行预设操作行为的知识点占有最后一条被  
执行预设操作行为的知识点的百分比；

25 统计每个相匹配的历史搜索记录中除去最后一条知识点的其他条被执行  
预设操作行为的知识点；

计算其中每一条被执行预设操作行为的知识点占有所有被执行预设操作行  
为的知识点的百分比；

30 其中，所述预设操作行为包括对所述历史搜索记录中的知识点的点击行  
为。

4、如权利要求 1 所述的知识库搜索方法，其特征在于，所述计算所述预  
设知识库中每一条知识点的权值系数，包括：

获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，  
35 得到每一条知识点中的文档分词；及

计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

5、如权利要求 2 所述的知识库搜索方法，其特征在于，所述计算所述预  
设知识库中每一条知识点的权值系数，包括：

获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

5 6、如权利要求 3 所述的知识库搜索方法，其特征在于，所述计算所述预设知识库中每一条知识点的权值系数，包括：

获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

10 计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

7、如权利要求 4-6 任一项所述的知识库搜索方法，其特征在于，所述预设的算法为：

$$\text{Weight} = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1+idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw;$$

其中：

15 Nd 为知识点的个数，i 为第几个分词，n 为共 n 个分词；

Lbw 为最后一次操作行为权重，是根据最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比预先设置的；

Bw 为操作行为权重，是根据每一条被执行预设操作行为的知识点占有被执行预设操作行为的知识点的百分比预先设置的；及

20 Tf、Idf、Tw、Fw、Kw 及 Norm 为每一条知识点以及其中的文档分词的权值系数，分别代表所述每个知识点中的文档分词的词频、反词频、词权重以及位置权重，以及每一个知识点的知识权重、长度归一化因子。

8、一种知识库搜索装置，其特征在于，所述装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的知识库搜索程序，所述知识库搜索程序被所述处理器执行时实现如下步骤：

25 获取所述知识库中与当前接收的关键字相匹配的历史搜索记录，并计算所述历史搜索记录中的知识点的历史操作行为数据；

计算所述知识库中每一条知识点的权值系数；

30 根据所述权值系数以及所述知识点的历史操作行为数据，利用预设的算法，计算利用所述当前接收的关键字搜索到的每一条知识点的权值，并根据所述权值，将利用所述当前接收的关键字搜索到所有知识点进行排序呈现给用户。

9、如权利要求 8 所述的知识库搜索装置，其特征在于，所述获取所述知识库中与当前接收的关键字相匹配的历史搜索记录，并计算所述历史搜索记录中的知识点的历史操作行为数据，包括：

接收用户输入到所述知识库的关键字，并对接收的关键字执行分词操作；

获取所有用户在所述知识库的历史搜索记录，将所述历史搜索记录的关键字执行分词操作，并与当前接收的关键字进行匹配，以查找到与当前接收

的关键词匹配的历史搜索记录；及

计算所述历史搜索记录中的知识点的历史操作行为数据。

10、如权利要求 9 所述的知识库搜索装置，其特征在于，统计每个相匹配的历史搜索记录中最后一条被执行预设操作行为的知识点；

5 计算所述每个最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比；

统计每个相匹配的历史搜索记录中除去最后一条知识点的其他条被执行预设操作行为的知识点；

10 计算其中每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比；

其中，所述预设操作行为包括对所述历史搜索记录中的知识点的点击行为。

11、如权利要求 8 所述的知识库搜索装置，其特征在于，所述计算所述预设知识库中每一条知识点的权值系数，包括：

15 获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

20 12、如权利要求 9 所述的知识库搜索装置，其特征在于，所述计算所述预设知识库中每一条知识点的权值系数，包括：

获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

25 13、如权利要求 10 所述的知识库搜索装置，其特征在于，所述计算所述预设知识库中每一条知识点的权值系数，包括：

获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

30 计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

14、如权利要求 11-13 任一项所述的知识库搜索装置，其特征在于，所述预设的算法为：

$$\text{Weight} = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1+idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw ;$$

其中：

35 Nd 为知识点的个数，i 为第几个分词，n 为共 n 个分词；

Lbw 为最后一次操作行为权重，是根据最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比预先设置的；

Bw 为操作行为权重，是根据每一条被执行预设操作行为的知识点占有所有

被执行预设操作行为的知识点的百分比预先设置的；及

Tf、Idf、Tw、Fw、Kw 及 Norm 为每一条知识点以及其中的文档分词的权值系数，分别代表所述每个知识点中的文档分词的词频、反词频、词权重以及位置权重，以及每一个知识点的知识权重、长度归一化因子。

- 5 15、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有知识库搜索程序，所述知识库搜索程序可被一个或者多个处理器执行，以实现如下步骤：

获取所述知识库中与当前接收的关键字相匹配的历史搜索记录，并计算所述历史搜索记录中的知识点的历史操作行为数据；

- 10 计算所述知识库中每一条知识点的权值系数；

根据所述权值系数以及所述知识点的历史操作行为数据，利用预设的算法，计算利用所述当前接收的关键字搜索到的每一条知识点的权值，并根据所述权值，将利用所述当前接收的关键字搜索到所有知识点进行排序呈现给用户。

- 15 16、如权利要求 15 所述的计算机可读存储介质，其特征在于，所述获取所述知识库中与当前接收的关键字相匹配的历史搜索记录，并计算所述历史搜索记录中的知识点的历史操作行为数据，包括：

接收用户输入到所述知识库的关键字，并对接收的关键字执行分词操作；

- 20 获取所有用户在所述知识库的历史搜索记录，将所述历史搜索记录的关键字执行分词操作，并与当前接收的关键字进行匹配，以查找到与当前接收的关键字匹配的历史搜索记录；及

计算所述历史搜索记录中的知识点的历史操作行为数据。

17、如权利要求 16 所述的计算机可读存储介质，其特征在于，统计每个相匹配的历史搜索记录中最后一条被执行预设操作行为的知识点；

- 25 计算所述每个最后一条被执行预设操作行为的知识点占有最后一条被执行预设操作行为的知识点的百分比；

统计每个相匹配的历史搜索记录中除去最后一条知识点的其他条被执行预设操作行为的知识点；

- 30 计算其中每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比；

其中，所述预设操作行为包括对所述历史搜索记录中的知识点的点击行为。

18、如权利要求 15 所述的计算机可读存储介质，其特征在于，所述计算所述预设知识库中每一条知识点的权值系数，包括：

- 35 获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

19、如权利要求 16 或 17 所述的计算机可读存储介质，其特征在于，所

述计算所述预设知识库中每一条知识点的权值系数，包括：

获取所述预设知识库中的所有知识点；

对每一条知识点中的所有内容按照预先确定的分词规则执行分词操作，得到每一条知识点中的文档分词；及

5 计算每一条知识点以及其中的文档分词的一种或者多种权值系数。

20、如权利要求 19 项所述的计算机可读存储介质，其特征在于，所述预设的算法为：

$$\text{Weight} = \sum_{i=0}^n Tw * \sqrt{Tf} * \left(1 + \log\left(\frac{Nd}{1+idf}\right)\right)^2 * Fw * Norm * Kw * Lbw * Bw ;$$

其中：

10 Nd 为知识点的个数，i 为第几个分词，n 为共 n 个分词；

Lbw 为最后一次操作行为权重，是根据最后一条被执行预设操作行为的知识点占有所有最后一条被执行预设操作行为的知识点的百分比预先设置的；

Bw 为操作行为权重，是根据每一条被执行预设操作行为的知识点占有所有被执行预设操作行为的知识点的百分比预先设置的；及

15 Tf、Idf、Tw、Fw、Kw 及 Norm 为每一条知识点以及其中的文档分词的权值系数，分别代表所述每个知识点中的文档分词的词频、反词频、词权重以及位置权重，以及每一个知识点的知识权重、长度归一化因子。

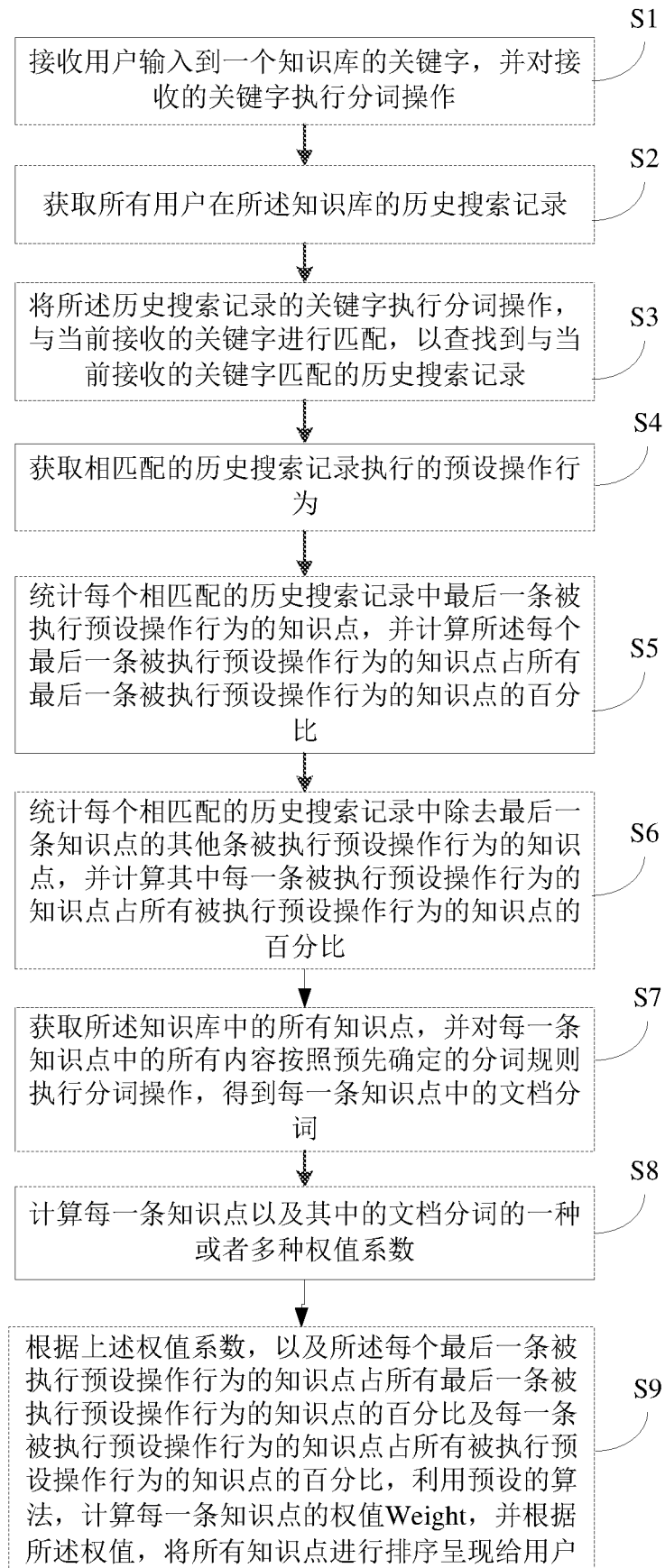


图 1

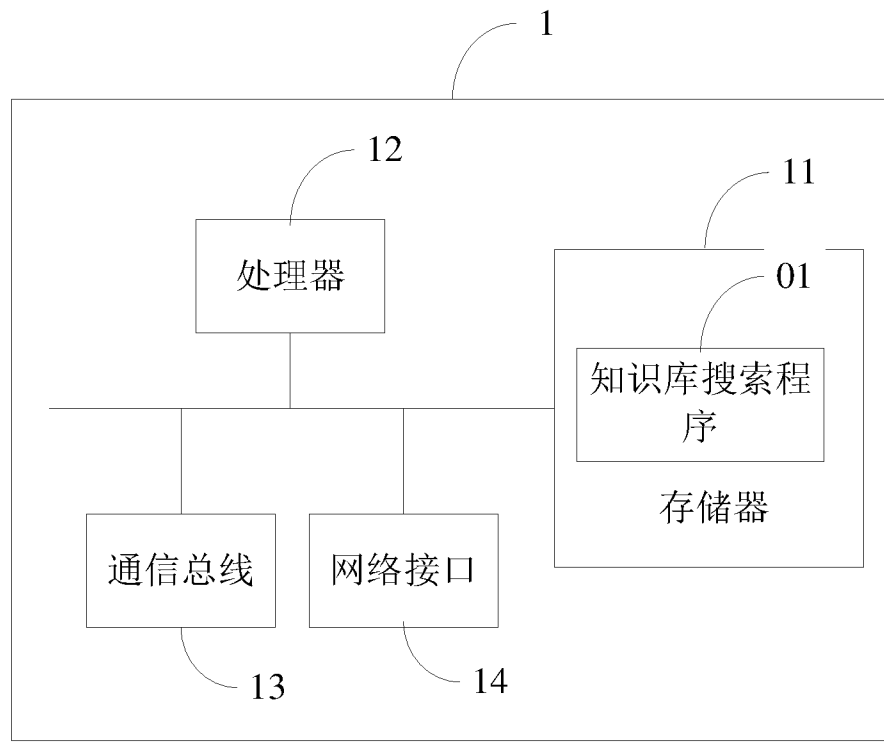


图 2

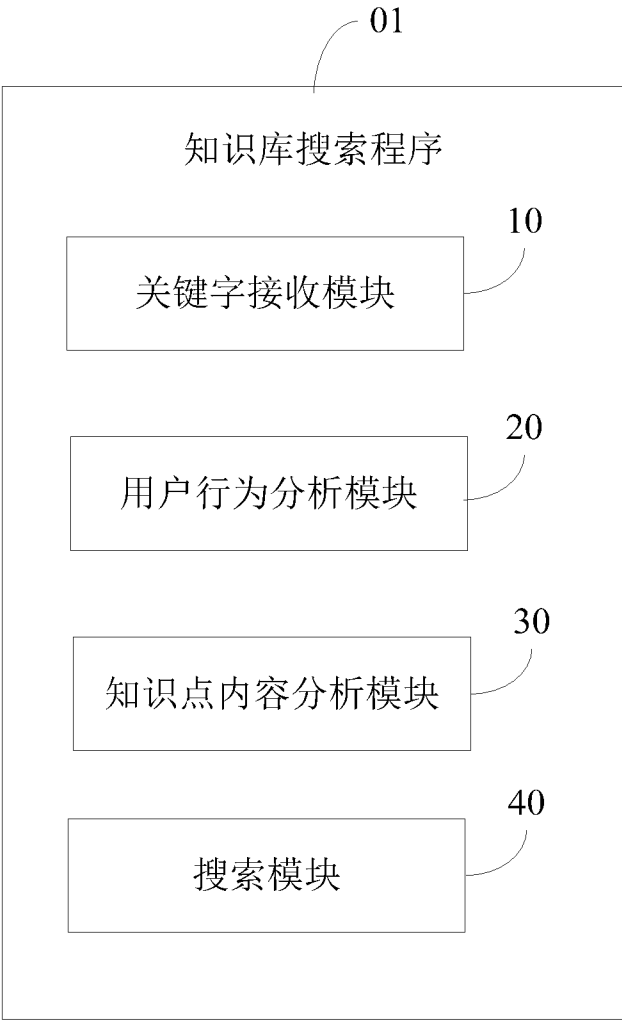


图 3

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/123597

**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 16/955(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 17/+, G06F 16/+

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; CNTXT; CNABS; CNKI: 搜索, 查询, 排序, 历史, 先前, 以前, 点击, 查看, 浏览, 操作, 选取, 记录, 条目, 项目, 知识点, search, query, order, sort, rank, history, foregone, previous, click, view, browse, operate, select, record, item, entry, knowledge point

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                        | Relevant to claim No.                    |
|-----------|-----------------------------------------------------------------------------------------------------------|------------------------------------------|
| X         | CN 104123332 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 29 October 2014 (2014-10-29)<br>claims 1-3 and 7-9 | 1, 2, 4, 5, 8, 9, 11, 12, 15, 16, 18, 19 |
| A         | CN 101996215 A (ALIBABA GROUP HOLDING LIMITED) 30 March 2011 (2011-03-30)<br>entire document              | 1-20                                     |
| A         | CN 102591969 A (BEIJING BAIDU NETWORK TECHNOLOGY CO., LTD.) 18 July 2012 (2012-07-18)<br>entire document  | 1-20                                     |
| A         | CN 108335137 A (BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD.) 27 July 2018 (2018-07-27)<br>entire document | 1-20                                     |
| A         | US 2018218087 A1 (WAL-MART STORES, INC.) 02 August 2018 (2018-08-02)<br>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

“&amp;” document member of the same patent family

Date of the actual completion of the international search

11 June 2019

Date of mailing of the international search report

21 June 2019

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/CN2018/123597**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| CN                                        | 104123332  | A  | 29 October 2014                      | WO                      | 2015109944 | A1 | 30 July 2015                         |
|                                           |            |    |                                      | CN                      | 104123332  | B  | 09 November 2018                     |
| CN                                        | 101996215  | A  | 30 March 2011                        | JP                      | 2013503391 | A  | 31 January 2013                      |
|                                           |            |    |                                      | EP                      | 2470987    | A4 | 21 September 2016                    |
|                                           |            |    |                                      | JP                      | 5596152    | B2 | 24 September 2014                    |
|                                           |            |    |                                      | US                      | 8346782    | B2 | 01 January 2013                      |
|                                           |            |    |                                      | US                      | 2012143816 | A1 | 07 June 2012                         |
|                                           |            |    |                                      | CN                      | 101996215  | B  | 24 July 2013                         |
|                                           |            |    |                                      | US                      | 8762391    | B2 | 24 June 2014                         |
|                                           |            |    |                                      | US                      | 2013086052 | A1 | 04 April 2013                        |
|                                           |            |    |                                      | WO                      | 2011025696 | A1 | 03 March 2011                        |
|                                           |            |    |                                      | EP                      | 2470987    | A1 | 04 July 2012                         |
|                                           |            |    |                                      | HK                      | 1149816    | A1 | 06 December 2013                     |
| CN                                        | 102591969  | A  | 18 July 2012                         | None                    |            |    |                                      |
| CN                                        | 108335137  | A  | 27 July 2018                         | None                    |            |    |                                      |
| US                                        | 2018218087 | A1 | 02 August 2018                       | None                    |            |    |                                      |

## 国际检索报告

国际申请号

PCT/CN2018/123597

| <b>A. 主题的分类</b><br>G06F 16/955 (2019.01) i<br><br>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>B. 检索领域</b><br>检索的最低限度文献 (标明分类系统和分类号)<br>G06F 17/+, G06F 16/+<br><br>包含在检索领域中的除最低限度文献以外的检索文献<br><br>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))<br>DWPI; SIPOABS; CNTXT; CNABS; CNKI; 搜索, 查询, 排序, 历史, 先前, 以前, 点击, 查看, 浏览, 操作, 选取, 记录, 条目, 项目, 知识点, search, query, order, sort, rank, history, foregone, previous, click, view, browse, operate, select, record, item, entry, knowledge point                                                                                                                                                                                                                                                                                                                                                                                      |                                                                           |                                              |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| <b>C. 相关文件</b> <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104123332 A (腾讯科技深圳有限公司) 2014年 10月 29日 (2014 - 10 - 29)<br/>权利要求1-3、7-9</td> <td>1-2, 4-5, 8-9, 11-12, 15-16, 18-19</td> </tr> <tr> <td>A</td> <td>CN 101996215 A (阿里巴巴集团控股有限公司) 2011年 3月 30日 (2011 - 03 - 30)<br/>全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 102591969 A (北京百度网讯科技有限公司) 2012年 7月 18日 (2012 - 07 - 18)<br/>全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108335137 A (北京三快在线科技有限公司) 2018年 7月 27日 (2018 - 07 - 27)<br/>全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2018218087 A1 (WAL MART STORES INC) 2018年 8月 2日 (2018 - 08 - 02)<br/>全文</td> <td>1-20</td> </tr> </tbody> </table> |                                                                           |                                              | 类 型* | 引用文件, 必要时, 指明相关段落 | 相关的权利要求 | X | CN 104123332 A (腾讯科技深圳有限公司) 2014年 10月 29日 (2014 - 10 - 29)<br>权利要求1-3、7-9 | 1-2, 4-5, 8-9, 11-12, 15-16, 18-19 | A | CN 101996215 A (阿里巴巴集团控股有限公司) 2011年 3月 30日 (2011 - 03 - 30)<br>全文 | 1-20 | A | CN 102591969 A (北京百度网讯科技有限公司) 2012年 7月 18日 (2012 - 07 - 18)<br>全文 | 1-20 | A | CN 108335137 A (北京三快在线科技有限公司) 2018年 7月 27日 (2018 - 07 - 27)<br>全文 | 1-20 | A | US 2018218087 A1 (WAL MART STORES INC) 2018年 8月 2日 (2018 - 08 - 02)<br>全文 | 1-20 |
| 类 型*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 引用文件, 必要时, 指明相关段落                                                         | 相关的权利要求                                      |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | CN 104123332 A (腾讯科技深圳有限公司) 2014年 10月 29日 (2014 - 10 - 29)<br>权利要求1-3、7-9 | 1-2, 4-5, 8-9, 11-12, 15-16, 18-19           |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | CN 101996215 A (阿里巴巴集团控股有限公司) 2011年 3月 30日 (2011 - 03 - 30)<br>全文         | 1-20                                         |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | CN 102591969 A (北京百度网讯科技有限公司) 2012年 7月 18日 (2012 - 07 - 18)<br>全文         | 1-20                                         |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | CN 108335137 A (北京三快在线科技有限公司) 2018年 7月 27日 (2018 - 07 - 27)<br>全文         | 1-20                                         |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| A                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | US 2018218087 A1 (WAL MART STORES INC) 2018年 8月 2日 (2018 - 08 - 02)<br>全文 | 1-20                                         |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| <input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                           |                                              |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| * 引用文件的具体类型:<br>“A” 认为不特别相关的表示了现有技术一般状态的文件<br>“E” 在国际申请日的当天或之后公布的在先申请或专利<br>“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)<br>“O” 涉及口头公开、使用、展览或其他方式公开的文件<br>“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<br>“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<br>“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<br>“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<br>“&” 同族专利的文件                                                                                                                                                                                                                                                                                                                                                                               |                                                                           |                                              |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| 国际检索实际完成的日期<br><br>2019年 6月 11日                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                           | 国际检索报告邮寄日期<br><br>2019年 6月 21日               |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |
| ISA/CN的名称和邮寄地址<br>中国国家知识产权局 (ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10) 62019451                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                           | 授权官员<br><br>吴广平<br><br>电话号码 (86-10) 62411829 |      |                   |         |   |                                                                           |                                    |   |                                                                   |      |   |                                                                   |      |   |                                                                   |      |   |                                                                           |      |

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2018/123597

| 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          | 104123332  | A  | 2014年 10月 29日  | WO   | 2015109944 | A1 | 2015年 7月 30日   |
|             |            |    |                | CN   | 104123332  | B  | 2018年 11月 9日   |
| CN          | 101996215  | A  | 2011年 3月 30日   | JP   | 2013503391 | A  | 2013年 1月 31日   |
|             |            |    |                | EP   | 2470987    | A4 | 2016年 9月 21日   |
|             |            |    |                | JP   | 5596152    | B2 | 2014年 9月 24日   |
|             |            |    |                | US   | 8346782    | B2 | 2013年 1月 1日    |
|             |            |    |                | US   | 2012143816 | A1 | 2012年 6月 7日    |
|             |            |    |                | CN   | 101996215  | B  | 2013年 7月 24日   |
|             |            |    |                | US   | 8762391    | B2 | 2014年 6月 24日   |
|             |            |    |                | US   | 2013086052 | A1 | 2013年 4月 4日    |
|             |            |    |                | WO   | 2011025696 | A1 | 2011年 3月 3日    |
|             |            |    |                | EP   | 2470987    | A1 | 2012年 7月 4日    |
|             |            |    |                | HK   | 1149816    | A1 | 2013年 12月 6日   |
| CN          | 102591969  | A  | 2012年 7月 18日   | 无    |            |    |                |
| CN          | 108335137  | A  | 2018年 7月 27日   | 无    |            |    |                |
| US          | 2018218087 | A1 | 2018年 8月 2日    | 无    |            |    |                |

表 PCT/ISA/210 (同族专利附件) (2015年1月)