

(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

WO 2020/056977 A1

(43) 国际公布日

2020 年 3 月 26 日 (26.03.2020)

(51) 国际专利分类号:

G06F 16/9535 (2019.01)

(21) 国际申请号:

PCT/CN2018/123595

(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:

201811096338.6 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)。 曲风龙(QU, FengLong); 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

(54) Title: KNOWLEDGE POINT PUSHING METHOD AND DEVICE, AND COMPUTER READABLE STORAGE MEDIUM

(54) 发明名称: 知识点推送方法、装置及计算机可读存储介质

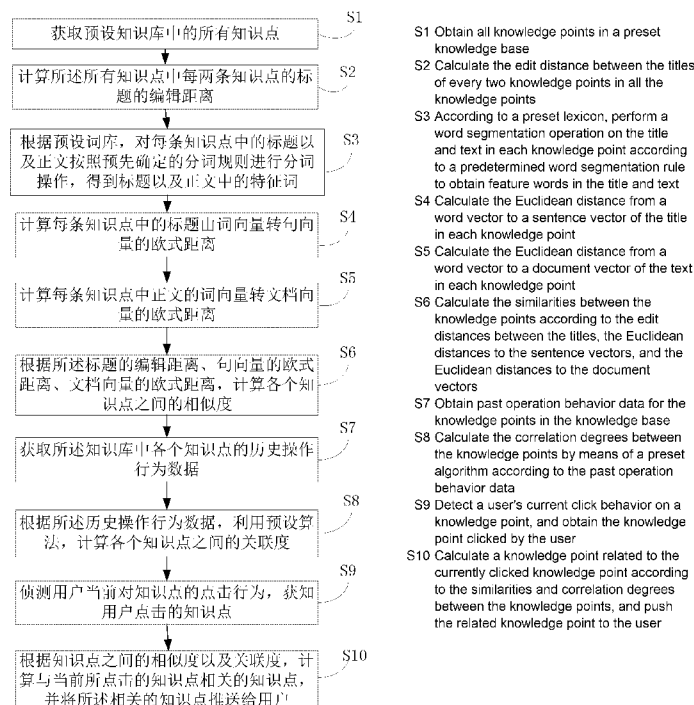


图 1

(57) Abstract: The present application relates to big data technology, and provides a knowledge point pushing method, comprising: obtaining all knowledge points in a preset knowledge base; calculating the similarities between the knowledge points according to the titles and the text of the knowledge points in the knowledge base; acquiring past operation behavior data for the knowledge points in the knowledge base, and calculating the correlation degrees between the knowledge points by means of a preset algorithm according to the past operation behavior data; and detecting a user's current click behavior on a knowledge point, obtaining the knowledge point clicked

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, 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))。

by the user, calculating a knowledge point related to the currently clicked knowledge point according to the similarities and correlation degrees between the knowledge points, and pushing the related knowledge point to the user. The present application also provides a knowledge point pushing device and a computer readable storage medium. According to the present application, when a user browses a knowledge point, other knowledge points more related to the knowledge point can be recommended to the user.

(57) 摘要: 本申请涉及一种大数据技术, 揭露了一种知识点推送方法, 包括: 获取预设知识库中的所有知识点; 根据所述知识库中各个知识点的标题以及正文内容, 计算各个知识点之间的相似度; 获取所述知识库中各个知识点的历史操作行为数据, 根据所述历史操作行为数据, 利用预设算法, 计算各个知识点之间的关联度; 及侦测用户当前对知识点的点击行为, 获知用户点击的知识点, 并根据知识点之间的相似度以及关联度, 计算与当前所点击的知识点相关的知识点, 将所述相关的知识点推送给用户。本申请还提出一种知识点推送装置以及一种计算机可读存储介质。本申请实现了在用户浏览知识点的时候, 能够给用户推荐与所述知识点更为相关的其他知识点。

知识点推送方法、装置及计算机可读存储介质

本申请要求于2018年9月19日提交中国专利局，申请号为201811096338.6、
发明名称为“知识点推送方法、装置及计算机可读存储介质”的中国专利申
5 请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及大数据技术领域，尤其涉及一种基于知识内容和用户行为协
同的知识点推送方法、装置及计算机可读存储介质。

10

背景技术

现有的通用搜索引擎都是基于关键词进行检索，在海量知识点中进行检
索时，检索结果大都是包含该关键词的网页，很难形成一个系统的、全面的、
关于该条目的详细介绍，从而也很难满足用户的知识需求。

15 知识推荐方法可以给用户提供更全面的知识体系。传统的知识推荐方法
通常是指根据用户输入的当前知识，从知识库中存储的现有知识中找出与用
户输入的知识在内容上相匹配的知识来推荐给用户。

这些传统的知识推荐方法大多是基于知识内容来进行知识匹配的。这种
方法仅采用类别作为知识是否相似的唯一匹配标准，可能导致知识内容的相
20 关度不高，匹配准确率差。利用这种方法推荐给用户的知识可能不是用户真
正感兴趣的知識。

发明内容

本申请提供一种知识点推送方法、装置及计算机可读存储介质，其主要
25 目的在于在用户浏览知识点的时候，能够给用户推荐与所述知识点更为相关
的其他知识点。

为实现上述目的，本申请提供的一种知识点推送方法，包括：

获取预设知识库中的所有知识点；

根据所述知识库中各个知识点的标题以及正文内容，计算各个知识点之

间的相似度;

获取所述知识库中各个知识点的历史操作行为数据, 根据所述历史操作行为数据, 利用预设算法, 计算各个知识点之间的关联度; 及

5 侦测用户当前对知识点的点击行为, 获知用户点击的知识点, 并根据知识点之间的相似度以及关联度, 计算与当前所点击的知识点相关的知识点, 将所述相关的知识点推送给用户。

此外, 为实现上述目的, 本申请还提供一种知识点推送装置, 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点推送程序, 所述知识点推送程序被所述处理器执行时实现如下步骤:

10 获取预设知识库中的所有知识点;

根据所述知识库中各个知识点的标题以及正文内容, 计算各个知识点之间的相似度;

获取所述知识库中各个知识点的历史操作行为数据, 根据所述历史操作行为数据, 利用预设算法, 计算各个知识点之间的关联度; 及

15 侦测用户当前对知识点的点击行为, 获知用户点击的知识点, 并根据知识点之间的相似度以及关联度, 计算与当前所点击的知识点相关的知识点, 将所述相关的知识点推送给用户。

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点推送程序, 所述知识点推送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点推送方法的步骤。

20 本申请提出的知识点推送方法、装置及计算机可读存储介质, 计算知识库中的各条知识点之间的相似度与关联度, 并在用户浏览所述知识库中的任何一条知识点时, 根据所述各条知识点之间的相似度与关联度, 计算与当前浏览的知识点相关的其他知识点, 并推送给用户。

25

附图说明

图1为本申请一实施例提供的知识点推送方法的流程示意图;

图2为本申请一实施例提供的知识点推送装置的内部结构示意图;

30 图3为本申请一实施例提供的知识点推送装置中知识点推送程序的模块示意图。

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。

具体实施方式

5 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点推送方法。参照图 1 所示，为本申请一实施例提供的知识点推送方法的流程示意图。该方法可以由一个装置执行，该装置可以由软件和/或硬件实现。

10 在本实施例中，知识点推送方法包括：

S1、获取预设知识库中的所有知识点。

本方案中，所述知识点可以是一个网页，一篇论文等。

本申请较佳实施例中，所述装置可以从与其相连接的知识库中获取的所有知识点。

15 S2、计算所述所有知识点中每两条知识点的标题的编辑距离。

所述编辑距离（Edit Distance），又称 Levenshtein 距离，是指两个字串之间，由一个转成另一个所需的最少编辑操作次数。许可的编辑操作包括将一个字符替换成另一个字符，插入一个字符，删除一个字符。一般来说，编辑距离越小，两个串的相似度越大。

20 例如：计算两个字符串 kitten 和 sitting 之间的编辑距离，需要包括如下转换：

1、kitten → sitten 将 k 替换成 s；

2、sitten → sittin 将 e 替换成 i；

3、sittin → sitting 添加 g。

25 本案较佳实施例可以设置每经过一次编辑，也就是变化（插入，删除，替换）需要花费的代价都是 1，因此，从 kitten 和 sitting 之间的编辑距离为 3。

S3、根据预设词库，对每条知识点中的标题以及正文按照预先确定的分词规则进行分词操作，得到标题以及正文中的特征词。

优选地，所述预先确定的分词规则如下：

30 按预设类型标点符号（例如，“，”、“。”、“！”、“；”等等）对获取的各

个标题以及正文进行短句拆分（例如，从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存的词库，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”）。

在本申请优选实施例中，还可以进一步对得到的特征词进行过滤处理，具体地，过滤处理采用以下任一种或两种方式：方式一：根据词性对特征词进行过滤，保留名词、动词以及形容词；方式二：根据频次对特征词进行过滤，保留频次大于频次阈值的特征词，其中，频次是指特征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。

S4、计算每条知识点中的标题由词向量转句向量的欧式距离。

在本申请较佳实施例中，步骤 S4 进一步包括：计算所述标题的词向量。

计算词向量的方式包括：将所述特征词输入第一向量模型，获取所述第一向量模型输出的特征词的词向量。其中，所述第一向量模型可以包括：word2vector 模型。

进一步地，在本申请较佳实施例中，步骤 S4 还包括：根据所述词向量构造所述标题的句向量。

构造所述标题的句向量的方式包括，但不限于，将所有词向量进行矢量叠加并取平均值，获取所述句向量，或者将所述标题的词向量*标题所有特征

词的反词频，再除以所述标题所有特征词的个数。

其中，所述欧式距离是指在二维和三维空间中两点之间的实际距离。

S5、计算每条知识点中正文的词向量转文档向量的欧式距离。

本申请实施例中，步骤 S5 进一步包括：计算所述正文的词向量以及文档
5 向量。

计算所述正文的词向量的方法可以参照上述计算所述标题的词向量的方法。

计算所述正文的文档向量的方法包括如下如下几种：

一、将所述正文中的特征词输入第二向量模型，获取第二向量模型输出
10 的特征词的词向量。其中，所述第二向量模型可以包括：Doc2Vec 模型或者叫做 paragraph2vec 模型；

二、将所述正文的词向量用矢量叠加并取平均值的方法计算所述正文的文档向量；

三、将所述正文的词向量*正文所有特征词的反词频，再除以所述正文所
15 有特征词的个数。

S6、根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度。

本申请较佳实施例中，所述知识点之间的相似度采用如下方法计算：

相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档
20 向量的欧式距离。

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2>权重 3。

S7、获取所述知识库中各个知识点的历史操作行为数据。

本案较佳实施例中，所述历史操作行为数据包括用户对所述知识点的点击操作。所述点击行为是指，用户通过点击开启了某一个知识点进行浏览。

本申请较佳实施例可以从一个用户历史行为数据库中获取每一个用户对
25 各个知识点的点击行为。

S8、根据所述历史操作行为数据，利用预设算法，计算各个知识点之间的关联度。

本申请优选实施例中，所述预设算法为 FPG 算法。

所述 FPG 算法又称 FP-Growth 算法，是一种关联分析算法，它采取如下
30

分治策略：将提供频繁项集的数据库压缩到一棵频繁模式树（FP-tree），但仍保留项集关联信息。

FP-growth 算法的过程如下：

首先，构造 FP 树。

- 5 1: 扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；
- 2: 定义 minsup=20%，即最小支持度（知识点最少点击的次数）为 2；
- 3: 按照点击行为的次数的降序排列重新排列知识点集。（如果出现小于 2 的物品则需要删除）；
- 4: 按照点击行为记录出现次数重新调整知识点清单；
- 10 5: 进行 FP 树的构建。

其次，从 FP 树中挖掘频繁项集，获取被用户大量点击的知识点，并根据所述频繁项计算各个知识点之间的关联度。

S9、侦测用户当前对知识点的点击行为，获知用户点击的知识点。

- 15 S10、根据知识点之间的相似度以及关联度，计算与当前所点击的知识点相关的知识点，并将所述相关的知识点推送给用户。

本申请优选实施例中，与所点击的知识点相关的知识点的计算方法如下：
权重 4*知识点之间的相似度+权重 5*知识点之间的关联度。

其中，权重 4 以及权重 5 为根据经验设定，且权重 4 > 权重 5。

- 20 发明还提供一种知识点推送装置。参照图 2 所示，为本申请一实施例提供的知识点推送装置的内部结构示意图。

在本实施例中，所述知识点推送装置 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。该知识点推送装置 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。

- 25 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点推送装置 1 的内部存储单元，例如该知识点推送装置 1 的硬盘。存储器 11 在另一些实施例中也可以是知识点推送装置 1 的外部存储设备，例如知识点推送装置 1
- 30

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 还可以既包括知识点推送装置 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点推送装置 1 的应用软件及各类数据,例如知识点推送程序 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,例如执行知识点推送程序 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,用于显示在知识点推送装置 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 以及知识点推送程序 01 的知识点推送装置 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点推送装置 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 中存储有知识点推送程序 01;处理器 12 执行存储器 11 中存储的知识点推送程序 01 时实现如下步骤:

步骤一、获取预设知识库中的所有知识点。

本方案中,所述知识点可以是一个网页,一篇论文等。

本申请较佳实施例中,所述装置可以从与其相连接的知识库中获取的所有知识点。

步骤二、计算所述所有知识点中每两条知识点的标题的编辑距离。

所述编辑距离(Edit Distance),又称 Levenshtein 距离,是指两个字串之

间，由一个转成另一个所需的最少编辑操作次数。许可的编辑操作包括将一个字符替换成另一个字符，插入一个字符，删除一个字符。一般来说，编辑距离越小，两个串的相似度越大。

例如：计算两个字符串 kitten 和 sitting 之间的编辑距离，需要包括如下

5 转换：

1、kitten → sitten 将 k 替换成 s；

2、sitten → sittin 将 e 替换成 i；

3、sittin → sitting 添加 g。

本案较佳实施例可以设置每经过一次编辑，也就是变化（插入，删除，
10 替换）需要花费的代价都是 1，因此，从 kitten 和 sitting 之间的编辑距离为 3。

步骤三、根据预设词库，对每条知识点中的标题以及正文按照预先确定的分词规则进行分词操作，得到标题以及正文中的特征词。

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
15 个标题以及正文进行短句拆分（例如，从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，
20 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存的词库，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
25 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”）。

在本申请优选实施例中，还可以进一步对得到的特征词进行过滤处理，
30 具体地，过滤处理采用以下任一种或两种方式：

方式一：根据词性对特征词进行过滤，保留名词、动词以及形容词；

方式二：根据频次对特征词进行过滤，保留频次大于频次阈值的特征词，其中，频次是指特征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，
5 如金融词库和产品词库等。

步骤四、计算每条知识点中的标题由词向量转句向量的欧式距离。

在本申请较佳实施例中，步骤 S4 进一步包括：

计算所述标题的词向量。

计算词向量的方式包括：将所述特征词输入第一向量模型，获取所述第一
10 向量模型输出的特征词的词向量。其中，所述第一向量模型可以包括：
word2vector 模型。

进一步地，在本申请较佳实施例中，步骤 S4 还包括：

根据所述词向量构造所述标题的句向量。

构造所述标题的句向量的方式包括，但不限于，将所有词向量进行矢量
15 叠加并取平均值，获取所述句向量，或者将所述标题的词向量*标题所有特征
词的反词频，再除以所述标题所有特征词的个数。

其中，所述欧式距离是指在二维和三维空间中两点之间的实际距离。

步骤五、计算每条知识点中正文的词向量转文档向量的欧式距离。

本申请实施例中，步骤 S5 进一步包括：

20 计算所述正文的词向量以及文档向量。

计算所述正文的词向量的方法可以参照上述计算所述标题的词向量的方
法。

计算所述正文的文档向量的方法包括如下如下几种：

一、将所述正文中的特征词输入第二向量模型，获取第二向量模型输出
25 的特征词的文档向量。其中，所述第二向量模型可以包括：Doc2Vec 模型 或
者叫做 paragraph2vec 模型；

二、将所述正文的词向量用矢量叠加并取平均值的方法计算所述正文的
文档向量；

三、将所述正文的词向量*正文所有特征词的反词频，再除以所述正文所
30 有特征词的个数。

步骤六、根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度。

本申请较佳实施例中，所述知识点之间的相似度采用如下方法计算：

相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离。

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2>权重 3。

步骤七、获取所述知识库中各个知识点的历史操作行为数据。

本案较佳实施例中，所述历史操作行为数据包括用户对所述知识点的点击操作。所述点击行为是指，用户通过点击开启了某一个知识点，以浏览该知识点的內容。

本申请较佳实施例可以从一个用户历史行为数据库中获取每一个用户对各个知识点的点击行为。

步骤八、根据所述历史操作行为数据，利用预设算法，计算各个知识点之间的关联度。

本申请优选实施例中，所述预设算法为 FPG 算法。

所述 FPG 算法又称 FP-Growth 算法，是一种关联分析算法，它采取如下分治策略：将提供频繁项集的数据库压缩到一棵频繁模式树（FP-tree），但仍保留项集关联信息。

FP-growth 算法的过程如下：

第一步：构造 FP 树。

1：扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；

2：定义 minsup=20%，即最小支持度（知识点最少点击的次数）为 2；

3：按照点击行为的次数的降序排列重新排列知识点集。（如果出现小于 2 的物品则需要删除）；

4：按照点击行为记录出现次数重新调整知识点清单；

5：进行 FP 树的构建。

第二步：从 FP 树中挖掘频繁项集，获取被用户大量点击的知识点，并根据所述频繁项计算各个知识点之间的关联度。

步骤九、侦测用户当前对知识点的点击行为，获知用户点击的知识点。

步骤十、根据知识点之间的相似度以及关联度，计算与当前所点击的知

识点相关的知识点，并将所述相关的知识点推送给用户。

本申请优选实施例中，与所点击的知识点相关的知识点的计算方法如下：
权重 4*知识点之间的相似度+权重 5*知识点之间的关联度。

其中，权重 4 以及权重 5 为根据经验设定，且权重 4 > 权重 5。

5

可选地，在其他实施例中，知识点推送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，用于描述知识点推送程序在知识点推送装置中的执行过程。

10

例如，参照图 3 所示，为本申请知识点推送装置一实施例中的知识点推送程序的程序模块示意图，该实施例中，所述知识点推送程序可以被分割为知识点获取模块 10、相似度计算模块 20、关联度计算模块 30 以及推送模块 40，示例性地：

15

所述知识点获取模块 10 用于：获取预设知识库中的所有知识点。

所述相似度计算模块 20 用于：计算所述所有知识点中每两条知识点的标题的编辑距离，计算每条知识点中的标题由词向量转句向量的欧式距离，以及计算每条知识点中正文的词向量转文档向量的欧式距离，并根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度。

20

所述关联度计算模块 30 用于：获取所述知识库中各个知识点的历史操作行为数据，根据所述历史操作行为数据，利用预设算法，计算各个知识点之间的关联度。

所述推送模块 40 用于：侦测用户当前对知识点的点击行为，获知用户点击的知识点，并根据知识点之间的相似度以及关联度，计算与当前所点击的知识点相关的知识点，将所述相关的知识点推送给用户。

25

上述知识点获取模块 10、相似度计算模块 20、关联度计算模块 30 以及推送模块 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。

30

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点推送程序，所述知识点推送程序可被一个或多个处理器执行，以实现如下操作：

获取预设知识库中的所有知识点；

- 5 计算所述所有知识点中每两条知识点的标题的编辑距离，计算每条知识点中的标题由词向量转句向量的欧式距离，以及计算每条知识点中正文的词向量转文档向量的欧式距离，并根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度；

- 10 获取所述知识库中各个知识点的历史操作行为数据，根据所述历史操作行为数据，利用预设算法，计算各个知识点之间的关联度；

 侦测用户当前对知识点的点击行为，获知用户点击的知识点，并根据知识点之间的相似度以及关联度，计算与当前所点击的知识点相关的知识点，将所述相关的知识点推送给用户。

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

- 20 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。

- 25 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，30 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、一种知识点推送方法，其特征在于，所述方法包括：

获取预设知识库中的所有知识点；

5 根据所述知识库中各个知识点的标题以及正文内容，计算各个知识点之间的相似度；

获取所述知识库中各个知识点的历史操作行为数据，根据所述历史操作行为数据，利用预设算法，计算各个知识点之间的关联度；及

10 侦测用户当前对知识点的点击行为，获知用户点击的知识点，并根据知识点之间的相似度以及关联度，计算与当前所点击的知识点相关的知识点，将所述相关的知识点推送给用户。

2、如权利要求 1 所述的知识点推送方法，其特征在于，所述根据所述知识库中各个知识点的标题以及正文内容，计算各个知识点之间的相似度，包括：

15 计算所述所有知识点中每两条知识点的标题的编辑距离；

计算每条知识点中的标题由词向量转句向量的欧式距离，以及计算每条知识点中正文的词向量转文档向量的欧式距离；及

根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度。

20 3、如权利要求 2 所述的知识点推送方法，其特征在于，所述计算每条知识点中的标题由词向量转句向量的欧式距离，以及计算每条知识点中正文的词向量转文档向量的欧式距离，包括：

根据预设词库，对每条知识点中的标题以及正文按照预先确定的分词规则进行分词操作，得到标题以及正文中的特征词；

25 将所述标题的特征词输入第一向量模型，获取所述第一向量模型输出的所述标题的词向量，根据所述词向量构造所述标题的句向量；

将所述正文的特征词输入所述第一向量模型，获取所述第一向量模型输出的所述正文的词向量，及将所述正文中的特征词输入第二向量模型，获取第二向量模型输出的正文的文档向量。

30 4、如权利要求 2 所述的知识点推送方法，其特征在于，根据所述标题的

编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度的方法如下：

相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离；

5 其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2 大于权重 3。

5、如权利要求 3 所述的知识点推送方法，其特征在于，根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度的方法如下：

10 相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离；

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2 大于权重 3。

6、如权利要求 4 所述的知识点推送方法，其特征在于，所述预设算法为 FP-Growth 算法，其过程如下：

第一步：构造 FP 树，包括：

- 15 1: 扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；
2: 定义最小支持度；
3: 按照点击行为的次数的降序排列重新排列知识点集，其中如果出现小于所述最小支持度的知识点进行删除；
4: 按照点击行为记录出现次数重新调整知识点清单；
20 5: 进行 FP 树的构建；

第二步：从 FP 树中挖掘频繁项集，获取被用户大量点击的知识点，并根据所述频繁项计算各个知识点之间的关联度。

7、如权利要求 5 所述的知识点推送方法，其特征在于，所述预设算法为 FP-Growth 算法，其过程如下：

25 第一步：构造 FP 树，包括：

- 1: 扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；
2: 定义最小支持度；
3: 按照点击行为的次数的降序排列重新排列知识点集，其中如果出现小于所述最小支持度的知识点进行删除；
30 4: 按照点击行为记录出现次数重新调整知识点清单；

5: 进行 FP 树的构建;

第二步: 从 FP 树中挖掘频繁项集, 获取被用户大量点击的知识点, 并根据所述频繁项计算各个知识点之间的关联度。

8、一种知识点推送装置, 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点推送程序, 所述知识点推送程序被所述处理器执行时实现如下步骤:

获取预设知识库中的所有知识点;

根据所述知识库中各个知识点的标题以及正文内容, 计算各个知识点之间的相似度;

10 获取所述知识库中各个知识点的历史操作行为数据, 根据所述历史操作行为数据, 利用预设算法, 计算各个知识点之间的关联度; 及

侦测用户当前对知识点的点击行为, 获知用户点击的知识点, 并根据知识点之间的相似度以及关联度, 计算与当前所点击的知识点相关的知识点, 将所述相关的知识点推送给用户。

15 9、如权利要求 8 所述的知识点推送装置, 其特征在于, 所述根据各个知识点的标题以及正文内容, 计算各个知识点之间的相似度, 包括:

计算所述所有知识点中每两条知识点的标题的编辑距离;

计算每条知识点中的标题由词向量转句向量的欧式距离, 以及计算每条知识点中正文的词向量转文档向量的欧式距离; 及

20 根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离, 计算各个知识点之间的相似度。

10、如权利要求 9 所述的知识点推送装置, 其特征在于, 所述计算每条知识点中的标题由词向量转句向量的欧式距离, 以及计算每条知识点中正文的词向量转文档向量的欧式距离, 包括:

25 根据预设词库, 对每条知识点中的标题以及正文按照预先确定的分词规则进行分词操作, 得到标题以及正文中的特征词;

将所述标题的特征词输入第一向量模型, 获取所述第一向量模型输出的所述标题的词向量, 根据所述词向量构造所述标题的句向量;

30 将所述正文的特征词输入所述第一向量模型, 获取所述第一向量模型输出的所述正文的词向量, 及将所述正文中的特征词输入第二向量模型, 获取

第二向量模型输出的正文的文档向量。

11、如权利要求 9 所述的知识点推送装置，其特征在于，根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度的方法如下：

5 相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离；

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2 大于权重 3。

12、如权利要求 10 所述的知识点推送装置，其特征在于，根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度的方法如下：

10 相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离；

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2 大于权重 3。

13、如权利要求 11 所述的知识点推送装置，其特征在于，所述预设算法为 FP-Growth 算法，其过程如下：

15 第一步：构造 FP 树，包括：

1: 扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；

2: 定义最小支持度；

20 3: 按照点击行为的次数的降序排列重新排列知识点集，其中如果出现小于所述最小支持度的知识点进行删除；

4: 按照点击行为记录出现次数重新调整知识点清单；

5: 进行 FP 树的构建；

第二步：从 FP 树中挖掘频繁项集，获取被用户大量点击的知识点，并根据所述频繁项计算各个知识点之间的关联度。

25 14、如权利要求 12 所述的知识点推送装置，其特征在于，所述预设算法为 FP-Growth 算法，其过程如下：

第一步：构造 FP 树，包括：

1: 扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；

2: 定义最小支持度；

30 3: 按照点击行为的次数的降序排列重新排列知识点集，其中如果出现

现小于所述最小支持度的知识点进行删除;

4: 按照点击行为记录出现次数重新调整知识点清单;

5: 进行 FP 树的构建;

5 第二步: 从 FP 树中挖掘频繁项集, 获取被用户大量点击的知识点, 并根据所述频繁项计算各个知识点之间的关联度。

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点推送程序, 所述知识点推送程序可被一个或者多个处理器执行, 以实现如下步骤:

获取预设知识库中的所有知识点;

10 根据所述知识库中各个知识点的标题以及正文内容, 计算各个知识点之间的相似度;

获取所述知识库中各个知识点的历史操作行为数据, 根据所述历史操作行为数据, 利用预设算法, 计算各个知识点之间的关联度; 及

15 侦测用户当前对知识点的点击行为, 获知用户点击的知识点, 并根据知识点之间的相似度以及关联度, 计算与当前所点击的知识点相关的知识点, 将所述相关的知识点推送给用户。

16、如权利要求 15 所述的计算机可读存储介质, 其特征在于, 所述根据各个知识点的标题以及正文内容, 计算各个知识点之间的相似度, 包括:

计算所述所有知识点中每两条知识点的标题的编辑距离;

20 计算每条知识点中的标题由词向量转句向量的欧式距离, 以及计算每条知识点中正文的词向量转文档向量的欧式距离; 及

根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离, 计算各个知识点之间的相似度。

25 17、如权利要求 16 所述的计算机可读存储介质, 其特征在于, 所述计算每条知识点中的标题由词向量转句向量的欧式距离, 以及计算每条知识点中正文的词向量转文档向量的欧式距离, 包括:

根据预设词库, 对每条知识点中的标题以及正文按照预先确定的分词规则进行分词操作, 得到标题以及正文中的特征词;

30 将所述标题的特征词输入第一向量模型, 获取所述第一向量模型输出的所述标题的词向量, 根据所述词向量构造所述标题的句向量;

将所述正文的特征词输入所述第一向量模型，获取所述第一向量模型输出的所述正文的词向量，及将所述正文中的特征词输入第二向量模型，获取第二向量模型输出的正文的文档向量。

18、如权利要求 16 所述的计算机可读存储介质，其特征在于，根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度的方法如下：

相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离；

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2 大于权重 3。

19、如权利要求 17 所述的计算机可读存储介质，其特征在于，根据所述标题的编辑距离、句向量的欧式距离、文档向量的欧式距离，计算各个知识点之间的相似度的方法如下：

相似度=权重 1*标题的编辑距离+权重 2*句向量的欧式距离+权重 3*文档向量的欧式距离；

其中，权重 1、权重 2 以及权重 3 为根据经验设定，且权重 2 大于权重 3。

20、如权利要求 18 或 19 所述的计算机可读存储介质，其特征在于，所述预设算法为 FP-Growth 算法，其过程如下：

第一步：构造 FP 树，包括：

1: 扫描用户历史操作行为数据库，获取每个知识点的点击行为记录；

2: 定义最小支持度；

3: 按照点击行为的次数的降序排列重新排列知识点集，其中如果出现小于所述最小支持度的知识点进行删除；

4: 按照点击行为记录出现次数重新调整知识点清单；

5: 进行 FP 树的构建；

第二步：从 FP 树中挖掘频繁项集，获取被用户大量点击的知识点，并根据所述频繁项计算各个知识点之间的关联度。

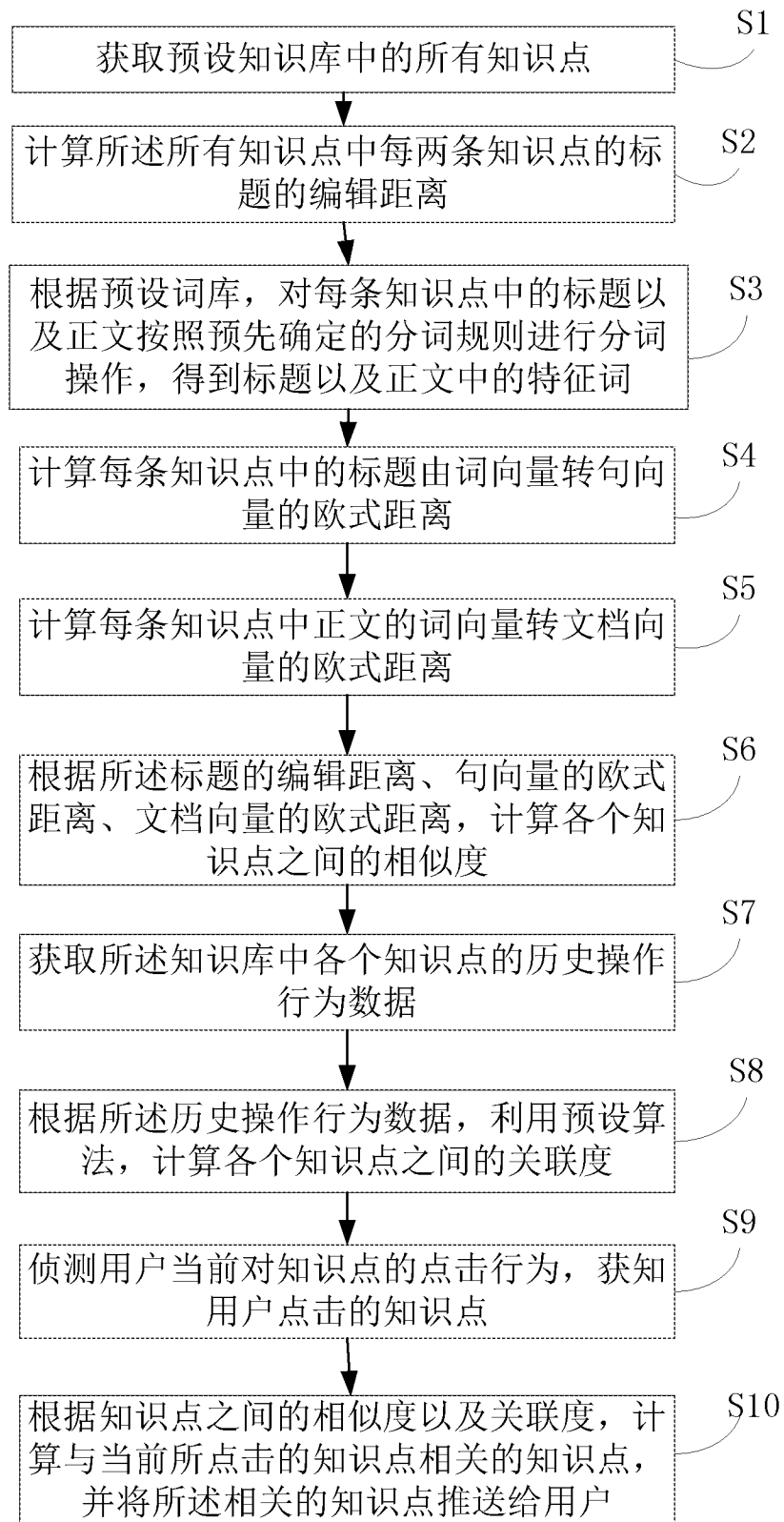


图 1

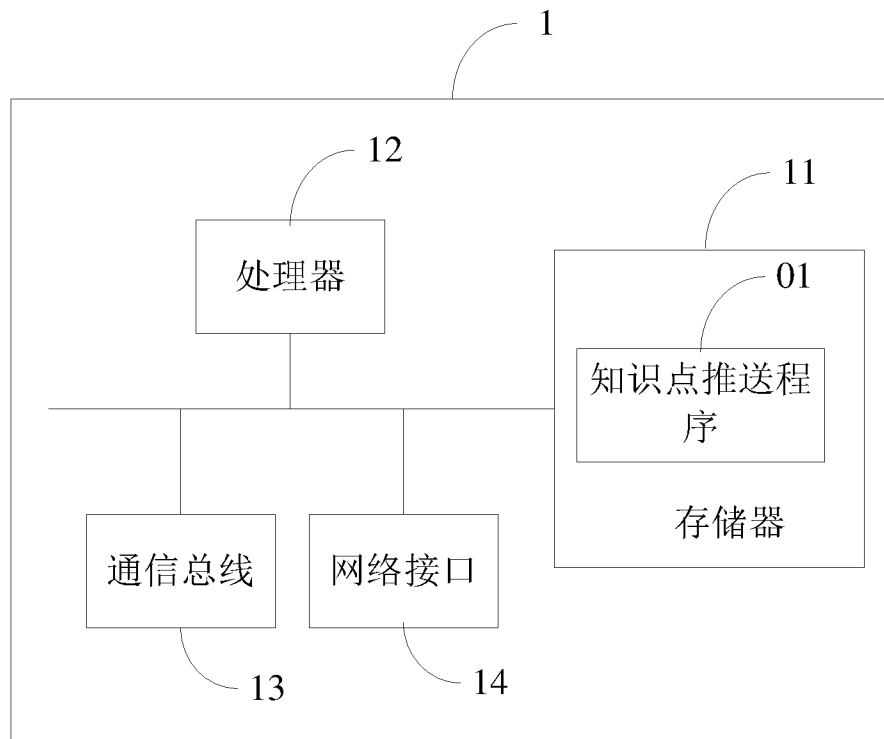


图 2

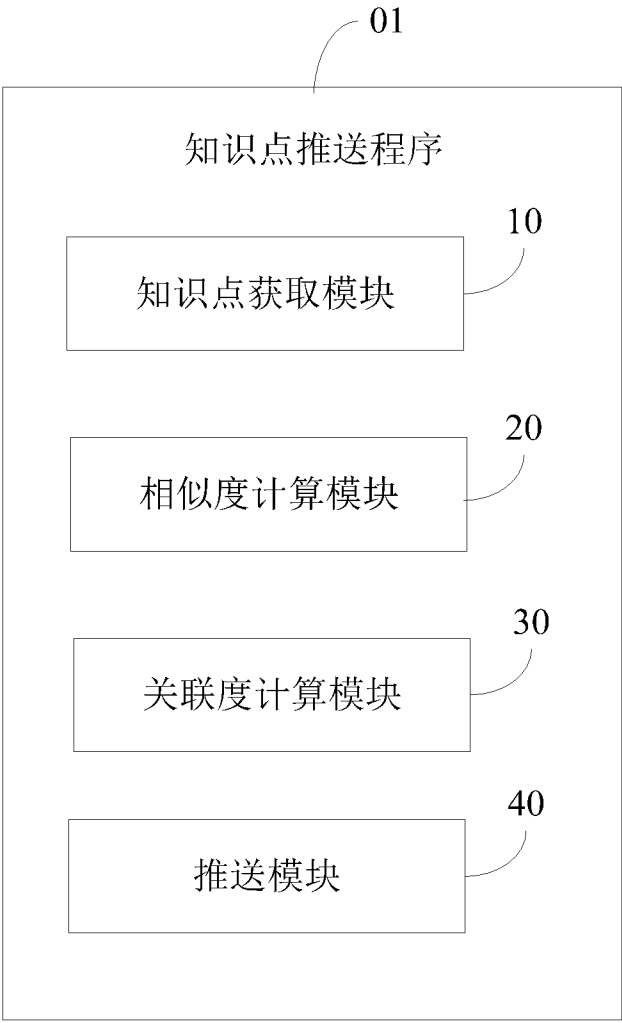


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/123595

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/9535(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)

CNABS; DWPI; CNKI; SIPOABS: 操作, 关键词, 点击, 关联度, 行为, 推送, 词语, 推荐, 知识点, recommendation, keyword, association

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106874308 A (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 20 June 2017 (2017-06-20) description, paragraphs [0153]-[0175], and figure 4	1-20
A	CN 104216874 A (GUANGXI COLLEGE OF EDUCATION) 17 December 2014 (2014-12-17) entire document	1-20
A	CN 102750334 A (INSTITUTE OF INFORMATION ON SCIENCE AND TECHNOLOGY OF AGRICULTURE, BEIJING ACADEMY OF AGRICULTURE AND FORESTRY SCIENCES ET AL.) 24 October 2012 (2012-10-24) 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

06 May 2019

Date of mailing of the international search report

13 May 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC (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/CN2018/123595

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106874308	A	20 June 2017	None			
CN	104216874	A	17 December 2014	CN	104216874	B	29 March 2017
CN	102750334	A	24 October 2012	CN	102750334	B	21 December 2016

国际检索报告

国际申请号

PCT/CN2018/123595

A. 主题的分类

G06F 16/9535 (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))

CNABS; DWPI; CNKI; SIPOABS: 操作, 关键词, 点击, 关联度, 行为, 推送, 词语, 推荐, 知识点, recommendation, keyword, association

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 106874308 A (北京搜狗科技发展有限公司) 2017年 6月 20日 (2017 - 06 - 20) 说明书[0153]-[0175]及图4	1-20
A	CN 104216874 A (广西教育学院) 2014年 12月 17日 (2014 - 12 - 17) 全文	1-20
A	CN 102750334 A (北京市农林科学院农业科技信息研究所, 等) 2012年 10月 24日 (2012 - 10 - 24) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 5月 6日

国际检索报告邮寄日期

2019年 5月 13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

受权官员

马驰

传真号 (86-10) 62019451

电话号码 62412168

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/123595

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	106874308	A	2017年 6月 20日	无			
CN	104216874	A	2014年 12月 17日	CN	104216874	B	2017年 3月 29日
CN	102750334	A	2012年 10月 24日	CN	102750334	B	2016年 12月 21日