(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112835923 A
(43)申请公布日 2021.05.25

(21)申请号 202110141804.3

(22)申请日 2021.02.02

(71)申请人 中国工商银行股份有限公司
地址 100140 北京市西城区复兴门内大街55号

(72)发明人 兰亭 徐琳玲 张闯 强锋

(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 任默闻 孙乳笋

(51)Int.Cl.
G06F 16/242(2019.01)
G06F 40/216(2020.01)
G06F 16/2458(2019.01)
G06F 16/248(2019.01)

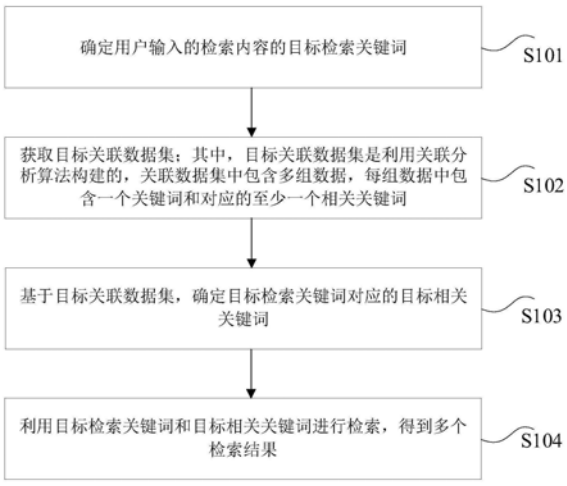
权利要求书3页 说明书10页 附图2页

(54)发明名称

一种相关检索方法、装置和设备

(57)摘要

本说明书实施例提供了一种相关检索方法、装置和设备，涉及大数据技术领域，其中，该方法包括：确定用户输入的检索内容的目标检索关键词；获取目标关联数据集；其中，所述目标关联数据集是利用关联分析算法构建的，所述关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词；基于所述目标关联数据集，确定所述目标检索关键词对应的目标相关关键词；利用所述目标检索关键词和所述目标相关关键词进行检索，得到多个检索结果。在本说明书实施例中，可以利用目标相关关键词检索到未包含目标检索关键词的相关内容，有效提高了检索结果的全面性，可以为用户更准确地查询到相关的检索结果，提高了用户体验感。



CN 112835923 A

1.一种相关检索方法,其特征在于,包括:

确定用户输入的检索内容的目标检索关键词;

获取目标关联数据集;其中,所述目标关联数据集是利用关联分析算法构建的,所述关联数据集中包含多组数据,每组数据中包含一个关键词和对应的至少一个相关关键词;

基于所述目标关联数据集,确定所述目标检索关键词对应的目标相关关键词;

利用所述目标检索关键词和所述目标相关关键词进行检索,得到多个检索结果。

2.根据权利要求1所述的方法,其特征在于,在获取目标关联数据集之前,还包括:

确定目标数据库中记录的各个内容对应的关键词;

建立所述各个内容与关键词之间的对应关系;

根据所述各个内容与关键词之间的对应关系,设置目标支持度;

利用关联分析算法,根据所述目标支持度构建各个关键词的频繁模式树;其中,所述频繁模式树中的每个节点表征一个关键词;

基于所述频繁模式树构建所述目标关联数据集。

3.根据权利要求2所述的方法,其特征在于,基于所述频繁模式树构建所述目标关联数据集,包括:

基于所述频繁模式树筛选出各个关键词的相关关键词;

建立各个关键词与相关关键词之间的对应关系,得到初始关联数据集;

获取相关词评分表;其中,所述相关词评分表用于表征任意两个关键词之间的相关度;

基于所述相关词评分表,对所述初始关联数据集进行优化处理,得到目标关联数据集;其中,所述优化处理包括添加相关关键词和删除相关关键词。

4.根据权利要求2所述的方法,其特征在于,确定目标数据库中记录的各个内容对应的关键词,包括:

在确定所述目标数据库中记录的目标内容有对应的关键词行的情况下,获取所述目标内容对应的关键词行;

对所述目标内容对应的关键词行进行预处理,得到所述目标内容对应的关键词;其中,所述预处理包括:根据分隔符拆分关键词行为多个关键词;

在确定所述目标数据库中记录的目标内容没有对应的关键词行的情况下,获取所述目标内容;

对所述目标内容进行预处理,得到所述目标内容对应的关键词;其中,所述预处理包括:分词和去停用词。

5.根据权利要求3所述的方法,其特征在于,基于所述相关词评分表,对所述初始关联数据集进行优化处理,得到目标关联数据集,包括:

基于所述相关词评分表,确定所述初始关联数据集中目标关键词对应的各个相关关键词的得分;

在目标关键词对应的第一相关关键词的得分小于等于第一预设阈值的情况下,删除所述第一相关关键词;

在所述相关词评分表中与所述目标关键词的相关度大于等于第二预设阈值的第二相关关键词在所述初始关联数据集中不存在的情况下,将所述第二相关关键词添加至与所述目标关键词对应的相关关键词中,得到所述目标关联数据集。

6.根据权利要求1所述的方法,其特征在于,在利用所述目标检索关键词和所述目标相关关键词进行检索,得到多个检索结果之后,还包括:

计算各个检索结果与所述目标检索关键词和所述目标相关关键词的相关程度;

根据所述各个检索结果与所述目标检索关键词和所述目标相关关键词的相关程度,对所述各个检索结果进行降序排列;

将降序排列后的各个检索结果展示给所述用户。

7.根据权利要求6所述的方法,其特征在于,按照以下公式计算各个检索结果与所述目标检索关键词和所述目标相关关键词的相关程度:

$$y=\frac{F\times\sum_{i=1}^{M}G_i}{H}$$

其中,y为检索结果与所述目标检索关键词和所述目标相关关键词的相关程度;F为在一个检索结果中出现的所述目标检索关键词和所述目标相关关键词的个数;H为检索结果的总字数;i为变量,i大于等于1小于等于M,M为所述目标检索关键词和所述目标相关关键词的总数;G_i为第i个关键词在一个检索结果中出现的次数;$\sum_{i=1}^{M}G_i$为所述目标检索关键词和所述目标相关关键词在一个检索结果中出现的总次数。

8.根据权利要求1所述的方法,其特征在于,获取目标关联数据集之前,还包括:

按照预设时间间隔统计目标搜索引擎中各个内容的点击量;

按照点击量对所述目标搜索引擎中各个内容进行排序,将排序前预设数量的内容作为目标内容;

确定各个目标内容的高频词,得到目标内容与高频词的对应关系;

获取相关词评分表;其中,所述相关词评分表中包含任意两个关键词的相关度的评分;

根据所述目标内容与高频词的对应关系,更新所述相关词评分表。

9.根据权利要求8所述的方法,其特征在于,根据所述目标内容与高频词的对应关系,更新所述相关词评分表,包括:

根据所述目标内容与高频词的对应关系,确定任意两个不同的高频词共现于同一内容的次数;

将共现于同一内容的次数大于等于第三预设阈值的两个第一高频词作为相关关键词,并将所述两个第一高频词在所述相关词评分表中相关度的评分加1;

在确定存在共现于同一内容的次数为0的两个第二高频词的情况下,确定所述两个第二高频词不共现于同一内容的次数;

在所述两个第二高频词不共现于同一内容的次数大于等于第四预设阈值的情况下,将所述两个第二高频词在所述相关词评分表中相关度的评分减1。

10.根据权利要求8所述的方法,其特征在于,确定各个目标内容的高频词,包括:

对所述各个目标内容进行预处理,得到所述各个目标内容中包含的多个词;其中,所述预处理包括:分词和去停用词;

将在同一目标内容中出现的频度大于等于第五预设阈值的词作为所述目标内容的高

频词,得到所述各个目标内容的高频词。

11.一种相关检索装置,其特征在于,包括:

第一确定模块,用于确定用户输入的检索内容的目标检索关键词;

获取模块,用于获取目标关联数据集;其中,所述目标关联数据集是利用关联分析算法构建的,所述关联数据集中包含多组数据,每组数据中包含一个关键词和对应的至少一个相关关键词;

第二确定模块,用于基于所述目标关联数据集,确定所述目标检索关键词对应的目标相关关键词;

检索模块,用于利用所述目标检索关键词和所述目标相关关键词进行检索,得到多个检索结果。

12.一种相关检索设备,其特征在于,包括处理器以及用于存储处理器可执行指令的存储器,所述处理器执行所述指令时实现权利要求1至10中任一项所述方法的步骤。

13.一种计算机可读存储介质,其特征在于,其上存储有计算机指令,所述指令被执行时实现权利要求1至10中任一项所述方法的步骤。

一种相关检索方法、装置和设备

技术领域

[0001] 本说明书实施例涉及大数据技术领域，特别涉及一种相关检索方法、装置和设备。

背景技术

[0002] 目前在大数据中进行检索时，主要是根据用户输入的内容的对关键词进行模糊匹配，但是采用这种方式进行检索容易遗漏未包含关键词的相关内容。因此，采用现有技术中的检索方案无法全面地检索到与用户输入的内容相关的信息。

[0003] 针对上述问题，目前尚未提出有效的解决方案。

发明内容

[0004] 本说明书实施例提供了一种相关检索方法、装置和设备，以解决现有技术中无法全面地检索到与用户输入的内容相关的信息的问题。

[0005] 本说明书实施例提供了一种相关检索方法，包括：确定用户输入的检索内容的目标检索关键词；获取目标关联数据集；其中，所述目标关联数据集是利用关联分析算法构建的，所述关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词；基于所述目标关联数据集，确定所述目标检索关键词对应的目标相关关键词；利用所述目标检索关键词和所述目标相关关键词进行检索，得到多个检索结果。

[0006] 本说明书实施例还提供了一种相关检索装置，包括：第一确定模块，用于确定用户输入的检索内容的目标检索关键词；获取模块，用于获取目标关联数据集；其中，所述目标关联数据集是利用关联分析算法构建的，所述关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词；第二确定模块，用于基于所述目标关联数据集，确定所述目标检索关键词对应的目标相关关键词；检索模块，用于利用所述目标检索关键词和所述目标相关关键词进行检索，得到多个检索结果。

[0007] 本说明书实施例还提供了一种相关检索设备，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器执行所述指令时实现所述相关检索方法的步骤。

[0008] 本说明书实施例还提供了一种计算机可读存储介质，其上存储有计算机指令，所述指令被执行时实现所述相关检索方法的步骤。

[0009] 本说明书实施例提供了一种相关检索方法，可以确定用户输入的检索内容的目标检索关键词，并获取目标关联数据集，其中，上述目标关联数据集是利用关联分析算法构建的。由于关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词，因此，可以基于目标关联数据集，确定出目标检索关键词对应的目标相关关键词。进一步的，可以利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果。从而可以利用目标相关关键词检索到未包含目标检索关键词的相关内容，在拓宽了对检索内容观察的视角的同时，有效提高了检索结果的全面性，可以为用户更准确地查询到相关的检索结果，提高了用户体验感。

附图说明

[0010] 此处所说明的附图用来提供对本说明书实施例的进一步理解，构成本说明书实施例的一部分，并不构成对本说明书实施例的限定。在附图中：

[0011] 图1是根据本说明书实施例提供的相关检索方法的步骤示意图；

[0012] 图2是根据本说明书实施例提供的相关检索装置的结构示意图；

[0013] 图3是根据本说明书实施例提供的相关检索设备的结构示意图。

具体实施方式

[0014] 下面将参考若干示例性实施方式来描述本说明书实施例的原理和精神。应当理解，给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本说明书实施例，而并非以任何方式限制本说明书实施例的范围。相反，提供这些实施方式是为了使本说明书实施例公开更加透彻和完整，并且能够将本公开的范围完整地传达给本领域的技术人员。

[0015] 本领域的技术人员知道，本说明书实施例的实施方式可以实现为一种系统、装置设备、方法或计算机程序产品。因此，本说明书实施例公开可以具体实现为以下形式，即：完全的硬件、完全的软件(包括固件、驻留软件、微代码等)，或者硬件和软件结合的形式。

[0016] 虽然下文描述流程包括以特定顺序出现的多个操作，但是应该清楚了解，这些过程可以包括更多或更少的操作，这些操作可以顺序执行或并行执行(例如使用并行处理器或多线程环境)。

[0017] 请参阅图1，本实施方式可以提供一种相关检索方法。该相关检索方法可以用于全面、准确地检索与用户输入的检索内容相关的信息。上述相关检索方法可以包括以下步骤。

[0018] S101：确定用户输入的检索内容的目标检索关键词。

[0019] 在本实施方式中，由于用户在进行检索时会在目标搜索引擎相应界面的输入框中输入想要检索的内容，因此，为了确定用户检索的意图，以及提高检索的有效性，可以先确定用户输入的检索内容的目标检索关键词。其中，目标检索关键词可以为一个也可以为多个，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0020] 在本实施方式中，用户输入的检索内容可以是一个或多个词，也可以是一句话，也可以是一段话，具体的可以根据实际情况确定，本说明书实施例对此不作限定。由于用户输入的检索内容可能会包含一些冗余信息或者用户输入的检索内容可能无法准确地表达用户的意图，因此，如果直接根据用户输入的检索内进行检索，则无法准确的进行检索。可以先确定检索内容的目标检索关键词，从而确定用户的检索意图。例如，用户输入的检索内容为北京的天气怎么样，目标检索关键词为：北京、天气，可以确定用户是想查询北京的天气情况，有效提高了检索的效率和准确性。

[0021] S102：获取目标关联数据集；其中，目标关联数据集是利用关联分析算法构建的，关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词。

[0022] 在本实施方式中，可以预先获取目标关联数据集，其中，上述目标关联数据集可以用于表征哪些关键词之间具有相关性，上述目标关联数据集可以是利用关联分析算法构建的，关联数据集中可以包含多组数据，每组数据中可以包含一个关键词和对应的至少一个相关关键词。

[0023] 在本实施方式中，上述目标关联数据集可以用于确定一个关键词是否存在相关关键词，以及存在哪些相关关键词。上述目标关联数据可以表格、文本或者图像等形式存储，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0024] 在本实施方式中，上述关联分析算法(FP-Growth)的分治策略为：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。频繁模式树是一种特殊的前缀树，由频繁项头表和项前缀树构成，关联分析算法可以基于频繁模式树的结构进行挖掘。

[0025] 在本实施方式中，获取目标关联数据集的方式可以包括：从预设数据库中拉取得到。当然可以理解的是，还可以采用其它可能的方式获取上述样本数据集，例如，可以按照预设路径查询得到，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0026] S103：基于目标关联数据集，确定目标检索关键词对应的目标相关关键词。

[0027] 在本实施方式中，可以基于上述目标关联数据集，确定出目标检索关键词对应的目标相关关键词。其中，目标检索关键词对应的目标相关关键词可以为一个也可以为多个，在一些情况下目标检索关键词也可以不存在相关关键词，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0028] 在本实施方式中，可以根据目标关联数据集中记录的关键词对应的相关关键词，确定目标检索关键词的目标相关关键词，从而可以确定出目标检索关键词其它的表述词汇或者与其相关的词汇。例如，西红柿的相关关键词为：番茄；苏州的相关关键词为：吴中区、工业园区、姑苏区、新区、相城区、吴江区等。

[0029] S104：利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果。

[0030] 在本实施方式中，为了提高检索的全面性和准确性，可以利用目标检索关键词和确定出的目标相关关键词同时进行检索，从而可以得到多个检索结果，并将多个检索结果展示给用户。相较于直接利用目标检索关键词进行检索的方式，本方案还可以检索到未包含关键词的相关内容，提高了检索结果的全面性和准确性。

[0031] 从以上的描述中，可以看出，本说明书实施例实现了如下技术效果：可以确定用户输入的检索内容的目标检索关键词，并获取目标关联数据集，其中，上述目标关联数据集是利用关联分析算法构建的。由于关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词，因此，可以基于目标关联数据集，确定出目标检索关键词对应的目标相关关键词。进一步的，可以利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果。从而可以利用目标相关关键词检索到未包含目标检索关键词的相关内容，在拓宽了对检索内容观察的视角的同时，有效提高了检索结果的全面性，可以为用户更准确地查询到相关的检索结果，提高了用户体验感。

[0032] 在一个实施方式中，在获取目标关联数据集之前，还可以包括：确定目标数据库中记录的各个内容对应的关键词，并建立各个内容与关键词之间的对应关系。可以根据各个内容与关键词之间的对应关系，设置目标支持度。并利用关联分析算法，根据目标支持度构建各个关键词的频繁模式树；其中，频繁模式树中的每个节点表征一个关键词。进一步的，可以基于频繁模式树构建目标关联数据集。

[0033] 在本实施方式中，可以利用历史存储的内容数据确定不同的关键词对应的相关关键词，从而构建目标关联数据集。上述目标数据库可以为搜索引擎的数据库，也可以为网站

或应用中用于存储内容数据的数据库，具体的可以根据应用场景确定，本说明书实施例对此不作限定。上述目标数据库中可以存储有供用户搜索的内容，例如，搜索引擎是用于查找文献的，则目标数据库中可以存储有各类文献；如果搜索引擎是用于查找美食的，则目标数据库中可以存储有各店铺名和店铺的菜单。

[0034] 在本实施方式中，各个内容与关键词之间的对应关系可以采用内容1：关键词1、关键词2；内容2：关键词1、关键词3的形式记录，当然还可以采用其它可能的形式记录，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0035] 在本实施方式中，支持度是关联分析算法的一种参数，支持度可以是某一关键词在内容与关键词之间的对应关系中出现的概率，可以用于描述该关键词的重要性。如果支持度偏小，可认为关键词本身出现的次数偏小，在数据量较大的情况下，筛除此类关键词可以减少运算量，提升性能。上述目标支持度可以是最小支持度，可以为用户关心的关联规则必须满足的最低重要性，只有满足最小支持度的关键词才能产生关联规则。上述目标支持度可以是根据各个内容与关键词之间的对应关系自定义的，具体的可以根据每个关键词出现的次数来确定，以防止目标支持度过大或过小。上述目标支持度可以为任意大于0的小数，例如：0.2、0.4等，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0036] 在本实施方式中，不满足目标支持度要求的将不会出现在最后的频繁模式树中，频繁模式树其通过链接来连接相似元素，被连起来的元素可以看成是一个链表。可以将每个内容的关键词按照支持度排序后，把每个内容的关键词按支持度降序依次插入到一棵以NULL(空)为根节点的树中，频繁模式树中的每个节点表征一个关键词，同时可以在每个节点处记录该节点出现的次数。

[0037] 在本实施方式中，关联分析算法的目的就是在多个出现的关键词中找到出现次数最多的关键词或者关键词集合，这里的最多指的是出现概率大于等于给定的阈值(目标支持度)。找到单个数据项的次数较为简单，只需要遍历计数即可，但是对于关键词的组合即关键词集的出现次数较难确定，比如，某个数据项A与数据项B的出现次数都是频繁的，但是他们的组合也就是说他们同时出现的次数却不频繁。数据集中出现较为频繁的关键词组合可以称为频繁项集，频繁项集中的关键词可以互为相关关键词。从而可以利用关联分析算法高效、准确地挖掘出存在关联的关键词，为确定目标检索关键词对应的目标相关关键词提供了数据基础。

[0038] 在一个实施方式中，确定目标数据库中记录的各个内容对应的关键词，可以包括：在确定目标数据库中记录的目标内容有对应的关键词行的情况下，获取目标内容对应的关键词行，并对目标内容对应的关键词行进行预处理，得到目标内容对应的关键词；其中，预处理包括：根据分隔符拆分关键词行为多个关键词。在确定目标数据库中记录的目标内容没有对应的关键词行的情况下，可以获取目标内容，并对目标内容进行预处理，得到目标内容对应的关键词；其中，预处理包括：分词和去停用词。

[0039] 在本实施方式中，可以预先确定目标数据库中是否记录有内容对应的关键词行，在确定有的情况下可以对关键词行进行根据分隔符拆分关键词行为多个关键词等预处理操作，从而得到内容对应的关键词。在确定没有的情况下可以直接获取内容，对内容进行分词、去停用词等预处理操作，从而得到内容对应的关键词。

[0040] 在本实施方式中，上述关键词行可以为内容对应的简介，上述关键词行可以用于

表征内容的核心思想。例如，在上述内容为某个书籍的情况下，关键词行则可以为该书籍的简介。当然，关键词行不限于上述举例，所属领域技术人员在本说明书实施例技术精髓的启示下，还可能做出其它变更，但只要其实现的功能和效果与本说明书实施例相同或相似，均应涵盖于本说明书实施例保护范围内。

[0041] 在一个实施方式中，基于频繁模式树构建目标关联数据集，可以包括：基于频繁模式树筛选出各个关键词的相关关键词，并建立各个关键词与相关关键词之间的对应关系，得到初始关联数据集。进一步的，可以获取相关词评分表；其中，相关词评分表用于表征任意两个关键词之间的相关度。并基于相关词评分表，对初始关联数据集进行优化处理，得到目标关联数据集；其中，优化处理包括添加相关关键词和删除相关关键词。

[0042] 在本实施方式中，可以基于频繁模式树筛选出各个关键词的相关关键词，假设频繁模式树中某一关键词为当前节点，其出现次数为B，可以设置参数上浮出现次数C和下浮出现次数D，在当前节点的前后节点中，筛选出现次数在区间[B-D,B+C]范围内的节点，根据符合筛选要求的关键词建立关键词-相关关键词的对应关系。例如：当前节点为m(关键词m)，其出现次数为B，若节点m的前X_1个节点中的节点m1、m2的出现次数在[B-D,B+C]范围内、节点m的后X_2个节点m3、m4的出现次数在[B-D,B+C]范围内，则节点m建立的关键词-相关关键词的对应关系为关键词m：相关关键词m1、相关关键词m2、相关关键词m3、相关关键词m4。可以按照上述步骤遍历频繁模式树中的各节点，生成各节点的关键词-相关关键词的对应关系。

[0043] 在本实施方式中，上述出现次数可以为关键词出现的次数，上述出现次数B可以为大于0的整数，上述上浮出现次数C和下浮出现次数D可以为大于0的整数，例如：2、3等，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0044] 在本实施方式中，上述初始关联数据集可以记录各个关键词的相关关键词，例如，以关键词1：相关关键词1、相关关键词2；关键词2：相关关键词1、相关关键词3的形式记录，上述初始关联数据集可以以表格、文本、图像等形式存储。当然，上述初始关联数据集不限于上述举例，所属领域技术人员在本说明书实施例技术精髓的启示下，还可能做出其它变更，但只要其实现的功能和效果与本说明书实施例相同或相似，均应涵盖于本说明书实施例保护范围内。

[0045] 在本实施方式中，由于利用关联分析算法挖掘得到的目标关联数据集可能会出现遗漏或者存在关联的不是很高的相关关键词的情况，因此，可以利用相关词评分表对目标关联数据集进行校正、优化。上述相关词评分表可以利用历史检索数据构建的，可以用于表征任意两个关键词之间的相关度，相关度越高则表示越有可能是相关的关键词。

[0046] 在本实施方式中，可以获取目标搜索引擎中各个内容的历史点击量，将点击量排序前预设数量的内容作为基础数据构建相关词评分表。可以确定各个内容中的高频词，将各个内容中的任意两个高频词作为一对数据记录下来，每对数据的初始分数为0。进一步的，可以统计任意两个不同的高频词共现于同一内容的次数，如果共现于同一内容的次数大于等于第三预设阈值则认为这两个第一高频词互为相关关键词，可以在将两个第一高频词在相关词评分表中相关度的评分加1。

[0047] 在本实施方式中，在两个第二高频词共现于同一内容的次数为0的前提下，可以统计这两个第二高频词不共现于同一内容的次数。在两个第二高频词不共现于同一内容的次

数大于等于第四预设阈值的情况下，可以将两个第二高频词在相关词评分表中相关度的评分减1。相关词评分表中评分越高，表明两个关键词的相关度越高，评分越小，则说明两个关键词的相关度越低。

[0048] 在本实施方式中，可以基于初始关联数据集中记录的两个相关的关键词在上述相关词评分表中的评分来确定是否需要保留该相关关键词。在一些情况下，在相关词评分表中关键词1和关键词2之间的相关度评分较高，但是初始关联数据集中关键词1的相关关键词中没有关键词2时，可以将关键词2添加至关键词1的相关关键词中。当然，优化处理的方式不限于上述举例，所属领域技术人员在本说明书实施例技术精髓的启示下，还可能做出其它变更，但只要其实现的功能和效果与本说明书实施例相同或相似，均应涵盖于本说明书实施例保护范围内

[0049] 在本实施方式中，可以利用相关词评分表对初始关联数据集进行进一步的优化，从而可以有效提高关联数据集中记录的关键词与相关关键词之间的对应关系的准确性，进而可以提高确定的目标检索关键词对应的目标相关关键词的准确性。

[0050] 在一个实施方式中，基于相关词评分表，对初始关联数据集进行优化处理，得到目标关联数据集，可以包括：基于相关词评分表，确定初始关联数据集中目标关键词对应的各个相关关键词的得分。在目标关键词对应的第一相关关键词存在上述相关词评分表中并且得分小于等于第一预设阈值的情况下，可以删除第一相关关键词。在相关词评分表中与目标关键词的相关度大于等于第二预设阈值的第二相关关键词在初始关联数据集中不存在的情况下，可以将第二相关关键词添加至与目标关键词对应的相关关键词中，得到目标关联数据集。

[0051] 在本实施方式中，可以预先设置第一预设阈值和第二预设阈值，上述第一预设阈值和第二预设阈值均为大于0的数值，上述第一预设阈值可以等于第二预设阈值，也可以不相等，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0052] 在本实施方式中，可以利用相关词评分表对初始关联数据集进行进一步的优化，从而可以有效提高关联数据集中记录的关键词与相关关键词之间的对应关系的准确性，进而可以提高确定的目标检索关键词对应的目标相关关键词的准确性。

[0053] 在一个实施方式中，在利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果之后，还可以包括：计算各个检索结果与目标检索关键词和目标相关关键词的相关程度，并根据各个检索结果与目标检索关键词和目标相关关键词的相关程度，对各个检索结果进行降序排列。进一步的，可以将降序排列后的各个检索结果展示给用户。

[0054] 在本实施方式中，由于不是每个检索结果都是用户想要的，即不同的检索结果与目标检索关键词和目标相关关键词的相关程度会存在差异。因此，可以按照各个检索结果与目标检索关键词和目标相关关键词的相关程度由高到低的顺序在用户的相应界面中展示，以便用户可以高效地找到与目标检索内容相关程度较高的检索结果，有效地提高了用户体验感。

[0055] 在一个实施方式中，可以按照以下公式计算各个检索结果与目标检索关键词和目标相关关键词的相关程度：

[0056] $$y=\frac{F\times\sum_{i=1}^{M}G_i}{H}$$

[0057] 其中，y为检索结果与目标检索关键词和目标相关关键词的相关程度；F为在一个检索结果中出现的目标检索关键词和目标相关关键词的个数；H为检索结果的总字数；i为变量，i大于等于1小于等于M，M为目标检索关键词和目标相关关键词的总数；G_i为第i个关键词在一个检索结果中出现的次数；$\sum_{i=1}^{M}G_i$为目标检索关键词和目标相关关键词在一个检索结果中出现的总次数。

[0058] 在本实施方式中，假设目标检索关键词和目标相关关键词包括：m1、m2、m3，M＝3，在检索结果1中m1的出现次数为1，m2的出现次数为2，m3的出现次数为0。则F＝2(一个是m1，一个是m2，由于m3出现次数为0，所以m3不统计在F内)；G＝1+2+0＝3(m1的出现次数1+m2的出现次数2+m3的出现次数0)。

[0059] 在一个实施方式中，在获取目标关联数据集之前，还可以包括：按照预设时间间隔统计目标搜索引擎中各个内容的点击量，并按照点击量对所述目标搜索引擎中各个内容进行排序，将排序前预设数量的内容作为目标内容。进一步的，可以确定各个目标内容的高频词，得到目标内容与高频词的对应关系。获取相关词评分表；其中，相关词评分表中包含任意两个关键词的相关度的评分，并根据目标内容与高频词的对应关系，更新相关词评分表。

[0060] 在本实施方式中，上述预设时间间隔可以为一天，也可以为一个星期，也可以为一个月，具体的可以根据实际情况确定，本说明书实施例对此不作限定。可以获取目标搜索引擎中在预设时间间隔内各个内容的点击量，例如，预设时间间隔为一天的情况下，可以统计当天从零点至24时目标搜索引擎中各个内容的点击量，当然，在一些实施例中也可以根据当天24时之前的目标搜索引擎中各个内容所有的历史点击数据，以统计各个内容的点击量。具体的，可以根据实际情况确定，本说明书实施例对此不作限定。

[0061] 在本实施方式中，由于用户点击量高的内容更具有参考性，因此，可以将排序前预设数量的内容作为目标内容。上述预设数量可以为大于0的整数，例如：10、20、36等，具体的可以根据实际情况确定，本说明书实施例对此不作限定。上述目标内容的高频词可以是在目标内容中出现频率较高的词，由于同一内容的高频词可能是互为相关的词，因此，可以根据目标内容与高频词的对应关系，更新相关词评分表。从而可以基于用户的反馈对相关词评分表进行不断的优化，有效提高了相关词评分表的准确度。

[0062] 在一个实施方式中，根据目标内容与高频词的对应关系，更新相关词评分表，可以包括：根据目标内容与高频词的对应关系，确定任意两个不同的高频词共现于同一内容的次数，可以将共现于同一内容的次数大于等于第三预设阈值的两个第一高频词作为相关关键词，并将两个第一高频词在相关词评分表中相关度的评分加1。进一步的，在确定存在共现于同一内容的次数为0的两个第二高频词的情况下，可以确定两个第二高频词不共现于同一内容的次数。在两个第二高频词不共现于同一内容的次数大于等于第四预设阈值的情况下，可以将两个第二高频词在相关词评分表中相关度的评分减1。

[0063] 在本实施方式中，可以预先设置第三预设阈值和第四预设阈值，上述第三预设阈值和第四预设阈值可以均为大于0的数值，上述第一预设阈值可以等于第二预设阈值，也可

以不相等，具体的可以根据实际情况确定，本说明书实施例对此不作限定。

[0064] 在本实施方式中，可以统计任意两个不同的高频词共现于同一内容的次数，如果共现于同一内容的次数大于等于第三预设阈值则认为这两个第一高频词互为相关关键词，可以在将两个第一高频词在相关词评分表中相关度的评分加1。

[0065] 在本实施方式中，在两个第二高频词共现于同一内容的次数为0的前提下，可以统计这两个第二高频词不共现于同一内容的次数。例如，目标内容包括：目标内容1：高频词1、高频词2；目标内容2：高频词1、高频词3。高频词1与高频词2共现过，所以不统计不共现次数；高频词1与高频词3共现过，所以不统计不共现次数；高频词1与高频词3共现次数为0，因此统计不共现于同一内容的次数为2。相关词评分表中评分越高，表明两个关键词的相关度越高，评分越小，则说明两个关键词的相关度越低。

[0066] 在本实施方式中，可以通过统计任意两个不同的高频词共现于同一内容的次数来确定任意两个高频词是否为相关的关键词，从而进一步对相关词评分表进行不断的优化，有效提高了相关词评分表的准确度。

[0067] 在一个实施方式中，确定各个目标内容的高频词，可以包括：对各个目标内容进行预处理，得到各个目标内容中包含的多个词；其中，预处理包括：分词和去停用词。进一步的，可以将在同一目标内容中出现的频度大于等于第五预设阈值的词作为目标内容的高频词，得到各个目标内容的高频词。

[0068] 在本实施方式中，可以通过分词工具将各个目标内容拆分为多个词，并通过去停用词对拆分得到的多个进行筛减，从而可以得到各个目标内容中包含的多个词。上述第五预设阈值可以为大于0的数值，例如，2、3等，具体的可以根据实际情况确定，本说明书实施例对此不作限定。在同一目标内容中出现的频度大于等于第五预设阈值的词可以认为是该目标内容的高频词，从而可以高效准确的生成目标内容与高频词之间的对应关系。

[0069] 基于同一发明构思，本说明书实施例中还提供了一种相关检索装置，如下面的实施例。由于相关检索装置解决问题的原理与相关检索方法相似，因此相关检索装置的实施可以参见相关检索方法的实施，重复之处不再赘述。以下所使用的，术语“单元”或者“模块”可以实现预定功能的软件和/或硬件的组合。尽管以下实施例所描述的装置较佳地以软件来实现，但是硬件，或者软件和硬件的组合的实现也是可能并被构想的。图2是本说明书实施例的相关检索装置的一种结构框图，如图2所示，可以包括：第一确定模块201、获取模块202、第二确定模块203、检索模块204，下面对该结构进行说明。

[0070] 第一确定模块201，可以用于确定用户输入的检索内容的目标检索关键词；

[0071] 获取模块202，可以用于获取目标关联数据集；其中，目标关联数据集是利用关联分析算法构建的，关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词；

[0072] 第二确定模块203，可以用于基于目标关联数据集，确定目标检索关键词对应的目标相关关键词；

[0073] 检索模块204，可以用于利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果。

[0074] 本说明书实施例实施方式还提供了一种电子设备，具体可以参阅图3所示的基于本说明书实施例提供的相关检索方法的电子设备组成结构示意图，电子设备具体可以包括

输入设备31、处理器32、存储器33。其中，输入设备31具体可以用于输入检索内容。处理器32具体可以用于确定用户输入的检索内容的目标检索关键词；获取目标关联数据集；其中，目标关联数据集是利用关联分析算法构建的，关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词；基于目标关联数据集，确定目标检索关键词对应的目标相关关键词；利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果。存储器33具体可以用于存储多个检索结果等参数。

[0075] 在本实施方式中，输入设备具体可以是用户和计算机系统之间进行信息交换的主要装置之一。输入设备可以包括键盘、鼠标、摄像头、扫描仪、光笔、手写输入板、语音输入装置等；输入设备用于把原始数据和处理这些数的程序输入到计算机中。输入设备还可以获取接收其他模块、单元、设备传输过来的数据。处理器可以按任何适当的方式实现。例如，处理器可以采取例如微处理器或处理器以及存储可由该(微)处理器执行的计算机可读程序代码(例如软件或固件)的计算机可读介质、逻辑门、开关、专用集成电路(Application Specific Integrated Circuit，ASIC)、可编程逻辑控制器和嵌入微控制器的形式等等。存储器具体可以是现代信息技术中用于保存信息的记忆设备。存储器可以包括多个层次，在数字系统中，只要能保存二进制数据的都可以是存储器；在集成电路中，一个没有实物形式的具有存储功能的电路也叫存储器，如RAM、FIFO等；在系统中，具有实物形式的存储设备也叫存储器，如内存条、TF卡等。

[0076] 在本实施方式中，该电子设备具体实现的功能和效果，可以与其它实施方式对照解释，在此不再赘述。

[0077] 本说明书实施例实施方式中还提供了一种基于相关检索方法的计算机存储介质，计算机存储介质存储有计算机程序指令，在计算机程序指令被执行时可以实现：确定用户输入的检索内容的目标检索关键词；获取目标关联数据集；其中，目标关联数据集是利用关联分析算法构建的，关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词；基于目标关联数据集，确定目标检索关键词对应的目标相关关键词；利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果。

[0078] 在本实施方式中，上述存储介质包括但不限于随机存取存储器(Random Access Memory，RAM)、只读存储器(Read-Only Memory，ROM)、缓存(Cache)、硬盘(Hard Disk Drive，HDD)或者存储卡(Memory Card)。所述存储器可以用于存储计算机程序指令。网络通信单元可以是依照通信协议规定的标准设置的，用于进行网络连接通信的接口。

[0079] 在本实施方式中，该计算机存储介质存储的程序指令具体实现的功能和效果，可以与其它实施方式对照解释，在此不再赘述。

[0080] 显然，本领域的技术人员应该明白，上述的本说明书实施例的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本说明书实施例不限制于任何特定的硬件和软件结合。

[0081] 虽然本说明书实施例提供了如上述实施例或流程图所述的方法操作步骤，但基于

常规或者无需创造性的劳动在所述方法中可以包括更多或者更少的操作步骤。在逻辑性上不存在必要因果关系的步骤中，这些步骤的执行顺序不限于本说明书实施例提供的执行顺序。所述的方法的在实际中的装置或终端产品执行时，可以按照实施例或者附图所示的方法顺序执行或者并行执行(例如并行处理器或者多线程处理的环境)。

[0082] 应该理解，以上描述是为了进行图示说明而不是为了进行限制。通过阅读上述描述，在所提供的示例之外的许多实施方式和许多应用对本领域技术人员来说都将是显而易见的。因此，本说明书实施例的范围不应该参照上述描述来确定，而是应该参照前述权利要求以及这些权利要求所拥有的等价物的全部范围来确定。

[0083] 以上所述仅为本说明书实施例的优选实施例而已，并不用于限制本说明书实施例，对于本领域的技术人员来说，本说明书实施例可以有各种更改和变化。凡在本说明书实施例的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本说明书实施例的保护范围之内。

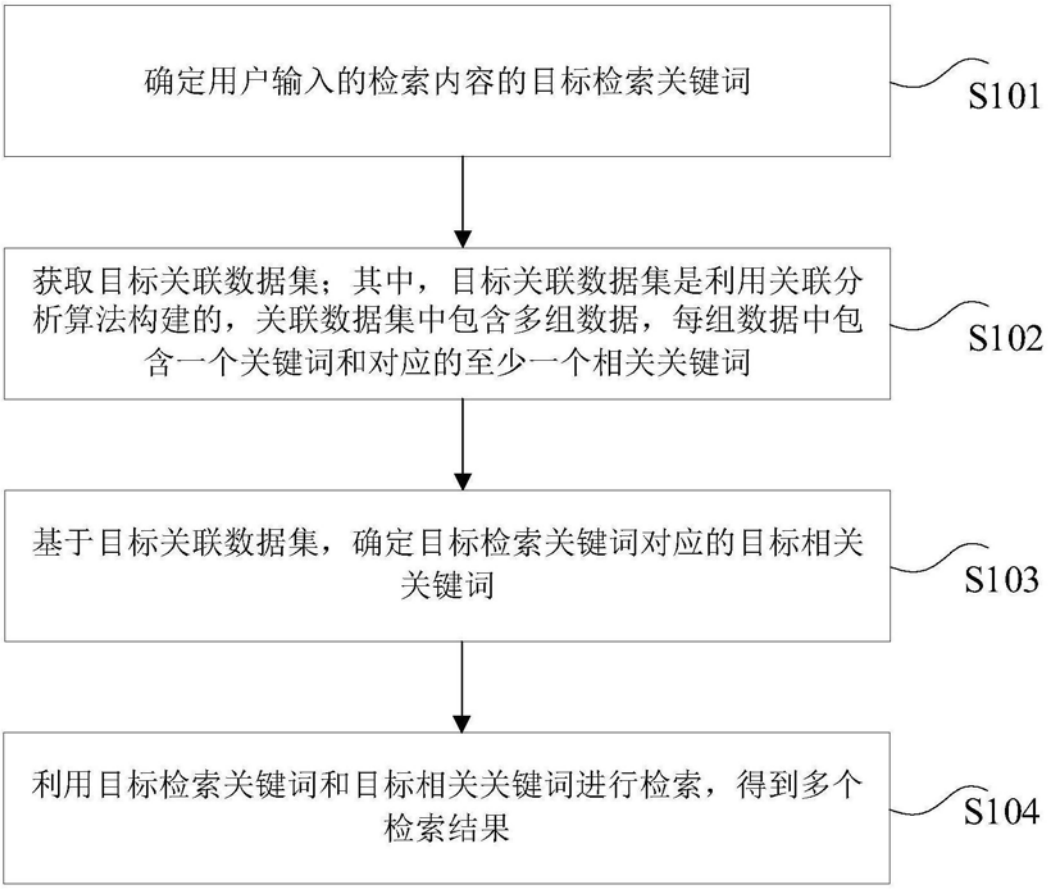

确定用户输入的检索内容的目标检索关键词
S101
获取目标关联数据集；其中，目标关联数据集是利用关联分析算法构建的，关联数据集中包含多组数据，每组数据中包含一个关键词和对应的至少一个相关关键词
S102
基于目标关联数据集，确定目标检索关键词对应的目标相关关键词
S103
利用目标检索关键词和目标相关关键词进行检索，得到多个检索结果
S104

图1


第一确定模块
201
获取模块
202
第二确定模块
203
检索模块
204

图2

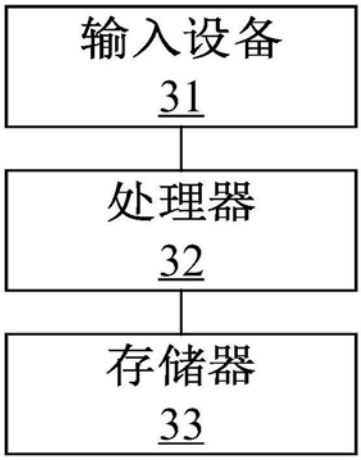


图3