



(12) 发明专利

(10) 授权公告号 CN 114491232 B

(45) 授权公告日 2023.03.24

(21) 申请号 202111602159.7

G06F 16/35 (2019.01)

(22) 申请日 2021.12.24

G06F 16/36 (2019.01)

G06F 40/166 (2020.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 114491232 A

(56) 对比文件

WO 2021212682 A1, 2021.10.28

(43) 申请公布日 2022.05.13

宁珊等.融合LSTM和LDA差异的新闻文本关键词抽取方法.《计算机工程与科学》.2020,(第01期),

(73) 专利权人 北京百度网讯科技有限公司

地址 100085 北京市海淀区上地十街10号

百度大厦二层

审查员 黄睿

(72) 发明人 盛广智 郑焯翰 蔡远俊 卞东海

黄焱晖 张记袁

(74) 专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

专利代理师 杜月

(51) Int. Cl.

G06F 16/9535 (2019.01)

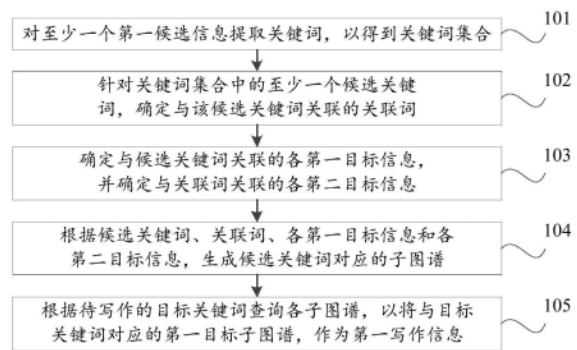
权利要求书3页 说明书18页 附图8页

(54) 发明名称

信息查询方法、装置、电子设备和存储介质

(57) 摘要

本公开提供了一种信息查询方法、装置、电子设备和存储介质,涉及知识图谱、智能搜索、深度学习等领域.具体实现方案为:对至少一个第一候选信息提取关键词,得到关键词集合,针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词;确定与候选关键词关联的各第一目标信息,及与关联词关联的各第二目标信息;根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱;根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱作为第一写作信息.由此,将与待写作的目标关键词对应的子图谱作为写作信息,从而向用户展示写作信息,可以便于用户进行思维发散与参考辅助写作。



1. 一种信息查询方法,所述方法包括:

对至少一个第一候选信息提取关键词,以得到关键词集合;

针对所述关键词集合中的至少一个候选关键词,确定与所述候选关键词关联的关联词;

确定与所述候选关键词关联的各第一目标信息,并确定与所述关联词关联的各第二目标信息;

根据所述候选关键词、关联词、各所述第一目标信息和各所述第二目标信息,生成所述候选关键词对应的子图谱;

根据待写作的目标关键词查询各子图谱,以将与所述目标关键词对应的第一目标子图谱,作为第一写作信息,其中,从各所述候选关键词中,确定与所述目标关键词匹配的候选关键词,将与所述目标关键词匹配的候选关键词所对应的子图谱,作为所述第一目标子图谱,并将所述第一目标子图谱作为所述第一写作信息;

其中,所述根据所述候选关键词、关联词、各所述第一目标信息和各所述第二目标信息,生成所述候选关键词对应的子图谱,包括:

将所述候选关键词作为对应的子图谱中的中心节点;

将所述关联词作为与所述中心节点关联的第一节点;

根据各所述第一目标信息的属性特征,确定与所述中心节点关联的第二节点;

根据各所述第二目标信息的属性特征,确定与所述第一节点关联的第三节点。

2. 根据权利要求1所述的方法,其中,所述属性特征,用于指示所述第一目标信息或所述第二目标信息所属的事件,所述生成所述候选关键词对应的子图谱之后,所述方法还包括:

根据待写作的目标事件查询各子图谱,以从各子图谱中确定第二目标子图谱,其中,所述第二目标子图谱中的第二节点的属性特征与所述目标事件匹配,所述第二节点的属性特征包括第一目标信息所属的事件;

将所述第二目标子图谱,作为第二写作信息。

3. 根据权利要求1所述的方法,其中,所述确定与所述候选关键词关联的关联词,包括:

对所述关键词集合中的任意两个关键词确定共同出现的次数,并根据所述次数,确定所述任意两个关键词的关联程度;

根据所述候选关键词与所述关键词集合中除所述候选关键词之外的其余关键词之间的关联程度,从所述其余关键词中确定与所述候选关键词关联的关联词。

4. 根据权利要求1所述的方法,其中,所述确定与所述候选关键词关联的各第一目标信息,包括:

确定与所述候选关键词匹配的多个第二候选信息;

从多个所述第二候选信息中,确定与所述候选关键词关联的各第一目标信息;其中,各所述第一目标信息之间的相似度未高于相似度阈值,且各所述第一目标信息属于同一事件的概率未高于概率阈值。

5. 根据权利要求4所述的方法,其中,所述确定与所述候选关键词匹配的多个第二候选信息,包括:

搜索发布时间与当前时间处于设定范围内的第三候选信息;

确定各所述第三候选信息与所述候选关键词之间的匹配度；

根据对各所述第三候选信息监听到的用户交互行为，确定各所述第三候选信息的热度；

根据各所述第三候选信息的匹配度和/或热度，从各所述第三候选信息中确定与所述候选关键词匹配的多个第二候选信息。

6. 根据权利要求1所述的方法，其中，所述确定与所述关联词关联的各第二目标信息，包括：

确定与所述关联词匹配的多个第四候选信息；

从多个所述第四候选信息中，确定与所述关联词关联的各第二目标信息，其中，各所述第二目标信息之间的相似度未高于相似度阈值，且各所述第二目标信息属于同一事件的概率未高于概率阈值。

7. 根据权利要求6所述的方法，其中，所述从多个所述第四候选信息中，确定与所述关联词关联的各第二目标信息，包括：

确定各所述第一目标信息所属的第一类别，并从各第一类别中确定目标类别，其中，属于所述目标类别的第一目标信息个数大于其他类别；

确定多个所述第四候选信息所属的第二类别，并从多个所述第四候选信息中，确定第二类别与所述目标类别相同的第五候选信息；

根据各所述第五候选信息，确定与所述关联词关联的各第二目标信息。

8. 一种信息查询装置，所述装置包括：

提取模块，用于对至少一个第一候选信息提取关键词，以得到关键词集合；

第一确定模块，用于针对所述关键词集合中的至少一个候选关键词，确定与所述候选关键词关联的关联词；

第二确定模块，用于确定与所述候选关键词关联的各第一目标信息；

第三确定模块，用于确定与所述关联词关联的各第二目标信息；

生成模块，用于根据所述候选关键词、关联词、各所述第一目标信息和各所述第二目标信息，生成所述候选关键词对应的子图谱；

处理模块，用于根据待写作的目标关键词查询各子图谱，以将与所述目标关键词对应的第一目标子图谱，作为第一写作信息，其中，从各所述候选关键词中，确定与所述目标关键词匹配的候选关键词，将与所述目标关键词匹配的候选关键词所对应的子图谱，作为所述第一目标子图谱，并将所述第一目标子图谱作为所述第一写作信息；

所述生成模块，具体用于：

将所述候选关键词作为对应的子图谱中的中心节点；

将所述关联词作为与所述中心节点关联的第一节点；

根据各所述第一目标信息的属性特征，确定与所述中心节点关联的第二节点；

根据各所述第二目标信息的属性特征，确定与所述第一节点关联的第三节点。

9. 根据权利要求8所述的装置，其中，所述属性特征，用于指示所述第一目标信息或所述第二目标信息所属的事件，所述处理模块，还用于：

根据待写作的目标事件查询各子图谱，以从各子图谱中确定第二目标子图谱，其中，所述第二目标子图谱中的第二节点的属性特征与所述目标事件匹配，所述第二节点的属性特

征包括第一目标信息所属的事件；

将所述第二目标子图谱,作为第二写作信息。

10. 根据权利要求8所述的装置,其中,所述第一确定模块,具体用于:

对所述关键词集合中的任意两个关键词确定共同出现的次数,并根据所述次数,确定所述任意两个关键词的关联程度;

根据所述候选关键词与所述关键词集合中除所述候选关键词之外的其余关键词之间的关联程度,从所述其余关键词中确定与所述候选关键词关联的关联词。

11. 根据权利要求8所述的装置,其中,所述第二确定模块,具体用于:

确定与所述候选关键词匹配的多个第二候选信息;

从多个所述第二候选信息中,确定与所述候选关键词关联的各第一目标信息;其中,各所述第一目标信息之间的相似度未高于相似度阈值,且各所述第一目标信息属于同一事件的概率未高于概率阈值。

12. 根据权利要求11所述的装置,其中,所述第二确定模块,具体用于:

搜索发布时间与当前时间处于设定范围内的第三候选信息;

确定各所述第三候选信息与所述候选关键词之间的匹配度;

根据对各所述第三候选信息监听到的用户交互行为,确定各所述第三候选信息的热度;

根据各所述第三候选信息的匹配度和/或热度,从各所述第三候选信息中确定与所述候选关键词匹配的多个第二候选信息。

13. 根据权利要求8所述的装置,其中,所述第三确定模块,具体用于:

确定与所述关联词匹配的多个第四候选信息;

从多个所述第四候选信息中,确定与所述关联词关联的各第二目标信息,其中,各所述第二目标信息之间的相似度未高于相似度阈值,且各所述第二目标信息属于同一事件的概率未高于概率阈值。

14. 根据权利要求13所述的装置,其中,所述第三确定模块,具体用于:

确定各所述第一目标信息所属的第一类别,并从各第一类别中确定目标类别,其中,属于所述目标类别的第一目标信息个数大于其他类别;

确定多个所述第四候选信息所属的第二类别,并从多个所述第四候选信息中,确定第二类别与所述目标类别相同的第五候选信息;

根据各所述第五候选信息,确定与所述关联词关联的各第二目标信息。

15. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-7中任一项所述的信息查询方法。

16. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-7中任一项所述的信息查询方法。

信息查询方法、装置、电子设备和存储介质

技术领域

[0001] 本公开涉及人工智能领域,具体涉及知识图谱、智能搜索、深度学习等技术领域,尤其涉及信息查询方法、装置、电子设备和存储介质。

背景技术

[0002] 目前,互联网已成为信息社会的基本通信工具,越来越多的用户通过互联网获取信息、发表看法、写作并发表文章等。比如,记者或编辑在新闻专题报道之前,会编写新闻文章,其中,媒体行业的新闻专题报道涉猎面较广,且形式多样。

[0003] 因此,为了提升用户的写作效率,如何提供更具智能和结构化的辅助能力来帮助用户进行写作,是非常重要的。

发明内容

[0004] 本公开提供了一种用于信息查询方法、装置、电子设备和存储介质。

[0005] 根据本公开的一方面,提供了一种信息查询方法,包括:

[0006] 对至少一个第一候选信息提取关键词,以得到关键词集合;

[0007] 针对所述关键词集合中的至少一个候选关键词,确定与所述候选关键词关联的关联词;

[0008] 确定与所述候选关键词关联的各第一目标信息,并确定与所述关联词关联的各第二目标信息;

[0009] 根据所述候选关键词、关联词、各所述第一目标信息和各所述第二目标信息,生成所述候选关键词对应的子图谱;

[0010] 根据待写作的目标关键词查询各子图谱,以将与所述目标关键词对应的第一目标子图谱,作为第一写作信息。

[0011] 根据本公开的另一方面,提供了一种信息查询装置,包括:

[0012] 提取模块,用于对至少一个第一候选信息提取关键词,以得到关键词集合;

[0013] 第一确定模块,用于针对所述关键词集合中的至少一个候选关键词,确定与所述候选关键词关联的关联词;

[0014] 第二确定模块,用于确定与所述候选关键词关联的各第一目标信息;

[0015] 第三确定模块,用于确定与所述关联词关联的各第二目标信息;

[0016] 生成模块,用于根据所述候选关键词、关联词、各所述第一目标信息和各所述第二目标信息,生成所述候选关键词对应的子图谱;

[0017] 处理模块,用于根据待写作的目标关键词查询各子图谱,以将与所述目标关键词对应的第一目标子图谱,作为第一写作信息。

[0018] 根据本公开的又一方面,提供了一种电子设备,包括:

[0019] 至少一个处理器;以及

[0020] 与所述至少一个处理器通信连接的存储器;其中,

[0021] 所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行本公开上述一方面提出的信息查询方法。

[0022] 根据本公开的再一方面,提供了一种计算机指令的非瞬时计算机可读存储介质,所述计算机指令用于使所述计算机执行本公开上述一方面提出的信息查询方法。

[0023] 根据本公开的还一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现本公开上述一方面提出的信息查询方法。

[0024] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其他特征将通过以下的说明书而变得容易理解。

附图说明

[0025] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0026] 图1为本公开实施例一所提供的信息查询方法的流程示意图;

[0027] 图2为本公开实施例二所提供的信息查询方法的流程示意图;

[0028] 图3为本公开实施例中的子图谱示意图;

[0029] 图4为本公开实施例三所提供的信息查询方法的流程示意图;

[0030] 图5为本公开实施例四所提供的信息查询方法的流程示意图;

[0031] 图6为本公开实施例五所提供的信息查询方法的流程示意图;

[0032] 图7为本公开实施例六所提供的信息查询方法的流程示意图;

[0033] 图8为本公开实施例七所提供的信息查询方法的流程示意图;

[0034] 图9为本公开实施例八所提供的信息查询方法的流程示意图;

[0035] 图10为本公开实施例九所提供的信息查询装置的结构示意图;

[0036] 图11示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。

具体实施方式

[0037] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0038] 随着互联网技术的不断发展,人工智能技术在媒体等多个行业逐渐地体现其应用价值,变得越来越重要。为了满足用户在写作前,对于发现写作内容之间的深度关联性、系统地了解一个新闻专题下的诸多方面内容的需求,需要设计一套基于互联网实时资讯的专题图谱构建系统。

[0039] 传统媒体行业的专题图谱,依赖于人工构建,即由人工检索后整理为专题图谱,极其耗时、耗力,产出效率很低,且图谱中的关联关系难以覆盖全面,即人工整理的方式,容易遗漏一些重要的关联关系。

[0040] 针对上述问题,本公开提出一种信息查询方法、装置、电子设备和存储介质。

[0041] 下面参考附图描述本公开实施例的信息查询方法、装置、电子设备和存储介质。

[0042] 图1为本公开实施例一所提供的信息查询方法的流程示意图。

[0043] 本公开实施例以该信息查询方法被配置于信息查询装置中来举例说明,该信息查询装置可以应用于任一电子设备中,以使该电子设备可以执行信息查询功能。

[0044] 其中,电子设备可以为任一具有计算能力的设备,例如可以为PC(Personal Computer,个人电脑)、移动终端、服务器等,移动终端例如可以为车载设备、手机、平板电脑、个人数字助理、穿戴式设备等具有各种操作系统、触摸屏和/或显示屏的硬件设备。

[0045] 如图1所示,该信息查询方法可以包括以下步骤:

[0046] 步骤101,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0047] 在本公开实施例中,第一候选信息可以为新闻资讯(或称为资讯稿件、新闻文本、新闻文档等),或者也可以为其他素材或其他信息,比如为音视频数据,或者为其他文本数据,比如用户发表的文章等,本公开对此并不做限制。

[0048] 在本公开实施例中,可以对至少一个第一候选信息提取关键词,根据提取的关键词,生成关键词集合。

[0049] 作为一种示例,以第一候选信息为新闻资讯进行示例,为了提升图谱中关联关系覆盖的全面性,可以对全网中每一篇新闻资讯进行关键词提取,比如可以采用wordrank算法,提取各篇新闻资讯中的关键词,得到关键词集合。

[0050] 步骤102,针对关键词集合中的至少一个候选关键词,确定与该候选关键词关联的关联词。

[0051] 在本公开实施例中,针对关键词集合中的至少一个候选关键词,可以确定与该候选关键词关联的关联词。其中,关联词的个数可以为至少一个,比如,为了提升图谱中关联关系覆盖的全面性,关联词的个数可以为多个。

[0052] 举例而言,以候选关键词为某电子产品的品牌进行示例,与该候选关键词关联的关联词可以为该品牌的总裁、该品牌下的其他产品、合作伙伴等等。

[0053] 步骤103,确定与候选关键词关联的各第一目标信息,并确定与关联词关联的各第二目标信息。

[0054] 在本公开实施例中,可以分别确定与候选关键词关联的各信息(比如素材),本公开中记为第一目标信息,以及确定与关联词关联的各信息,本公开中记为第二目标信息。

[0055] 步骤104,根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0056] 在本公开实施例中,可以根据候选关键词和关联词之间的关联关系、候选关键词与各第一目标信息之间的关联关系,以及关联词与各第二目标信息之间的关联关系,生成该候选关键词对应的子图谱。

[0057] 步骤105,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0058] 在本公开实施例中,待写作的目标关键词是根据用户的写作需求确定的,比如目标关键词可以为用户根据写作需求输入的。其中,输入方式包括但不限于触摸输入(如滑动、点击等)、键盘输入、语音输入等。例如,以用户为记者或新闻编辑工作者进行示例,目标关键词可以为近期发生的热门事件对应的关键词。

[0059] 在本公开实施例中,可以根据待写作的目标关键词,查询各候选关键词对应的子图谱,以从各候选关键词对应的子图谱中,确定与目标关键词对应的子图谱,本公开中记为

第一目标子图谱,即,可以从各候选关键词中,确定与目标关键词匹配的候选关键词,将与目标关键词匹配的候选关键词所对应的子图谱,作为第一目标子图谱。并将第一目标子图谱作为第一写作信息,从而向用户展示该第一写作信息,可以便于用户进行思维发散与参考辅助写作。并且,无需用户逐个查看并筛选信息,可以提升用户的写作效率。

[0060] 本公开实施例的信息查询方法,通过对至少一个第一候选信息提取关键词,以得到关键词集合,针对关键词集合中的至少一个候选关键词,确定与该候选关键词关联的关联词,并确定与候选关键词关联的各第一目标信息,及与关联词关联的各第二目标信息;根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱;根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。由此,通过根据各关键词之间的关联关系,关键词与信息(比如素材)之间的关联关系,自动生成各关键词对应的子图谱,而无需人工整理的方式得到,可以提升图谱的生成效率,降低人工成本。并且,将与待写作的目标关键词对应的第一目标子图谱,作为第一写作信息,从而向用户展示该第一写作信息,可以便于用户进行思维发散与参考辅助写作。此外,无需用户逐个查看并筛选信息,可以提升用户的写作效率。

[0061] 需要说明的是,本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均在征得用户同意的前提下进行,并且均符合相关法律法规的规定,且不违背公序良俗。

[0062] 为了清楚说明本公开上述实施例中是如何生成各候选关键词对应的子图谱的,本公开还提出一种信息查询方法。

[0063] 图2为本公开实施例二所提供的信息查询方法的流程示意图。

[0064] 如图2所示,该信息查询方法可以包括以下步骤:

[0065] 步骤201,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0066] 步骤202,针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词。

[0067] 步骤203,确定与候选关键词关联的各第一目标信息,并确定与关联词关联的各第二目标信息。

[0068] 步骤201至203的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0069] 步骤204,将候选关键词作为对应的子图谱中的中心节点。

[0070] 在本公开实施例中,可以将候选关键词,作为该候选关键词对应的子图谱中的中心节点。

[0071] 步骤205,将关联词作为与中心节点关联的第一节点。

[0072] 在本公开实施例中,可以将与候选关键词关联的关联词,作为与中心节点关联的第一节点。

[0073] 步骤206,根据各第一目标信息的属性特征,确定与中心节点关联的第二节点。

[0074] 在本公开实施例中,第一目标信息的属性特征可以包括第一目标信息的标题、第一目标信息的链接、第一目标信息所属的事件、第一目标信息中的关键词等等。

[0075] 在本公开实施例中,可以根据各第一目标信息的属性特征,确定与中心节点关联的第二节点。比如,可以将各第一目标信息的属性特征,作为与中心节点关联的各第二节点。

点。

[0076] 步骤207,根据各第二目标信息的属性特征,确定与第一节点关联的第三节点。

[0077] 同样地,第二目标信息的属性特征可以包括第二目标信息的标题、第二目标信息的链接、第二目标信息所属的事件、第二目标信息中的关键词等等。

[0078] 在本公开实施例中,可以根据各第二目标信息的属性特征,确定与第一节点关联的第三节点。比如,可以将各第二目标信息的属性特征,作为与第一节点关联的各第三节点。

[0079] 作为一种示例,以属性特征为信息的标题进行示例,假设候选关键词为A公司,则该A公司对应的子图谱可以如图3所示。

[0080] 步骤208,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0081] 步骤208的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0082] 本公开实施例的信息查询方法,通过将关联词作为与中心节点关联的第一节点;将关联词作为与中心节点关联的第一节点;根据各第二目标信息的属性特征,确定与第一节点关联的第三节点。由此,可以实现根据各关键词之间的关联关系,关键词与信息之间的关联关系,自动且有效地生成各关键词对应的子图谱。并且,还可以避免人工整理图谱的方式,存在遗漏关联关系的情况,提升关键词对应的子图谱生成的准确性和可靠性。

[0083] 在本公开实施例的一种可能的实现方式中,为了提升该方法的灵活性和适用性,以及满足用户实际的写作需求,还可以根据待写作的目标事件查询各子图谱,确定与目标事件匹配的子图谱,并作为写作信息。下面结合图4,对上述过程进行详细说明。

[0084] 图4为本公开实施例三所提供的信息查询方法的流程示意图。

[0085] 如图4所示,该信息查询方法可以包括以下步骤:

[0086] 步骤401,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0087] 步骤402,针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词。

[0088] 步骤403,确定与候选关键词关联的各第一目标信息,并确定与关联词关联的各第二目标信息。

[0089] 步骤404,将候选关键词作为对应的子图谱中的中心节点。

[0090] 步骤405,将关联词作为与中心节点关联的第一节点。

[0091] 步骤406,根据各第一目标信息的属性特征,确定与中心节点关联的第二节点。

[0092] 步骤407,根据各第二目标信息的属性特征,确定与第一节点关联的第三节点。

[0093] 步骤401至407的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0094] 步骤408,根据待写作的目标事件查询各子图谱,以从各子图谱中确定第二目标子图谱,其中,第二目标子图谱中的第二节点的属性特征与目标事件匹配,第二节点的属性特征包括第一目标信息所属的事件。

[0095] 在本公开实施例中,待写作的目标事件是根据用户的写作需求确定的,比如目标事件可以为用户根据写作需求输入的。其中,输入方式包括但不限于触摸输入(如滑动、点击等)、键盘输入、语音输入等。例如,以用户为记者或新闻编辑工作者进行示例,目标事件

可以为近期发生的热门事件。

[0096] 在本公开实施例中,可以根据待写作的目标事件,查询各候选关键词对应的子图谱,以从各候选关键词对应的子图谱中,确定与目标事件匹配的子图谱,本公开中记为第二目标子图谱,即,可以从各候选关键词对应的子图谱中,确定第二节点的属性特征与目标事件匹配的子图谱,并作为第二目标子图谱。其中,第二节点的属性特征包括第一目标信息所属的事件。

[0097] 步骤409,将第二目标子图谱,作为第二写作信息。

[0098] 在本公开实施例中,可以将第二目标子图谱作为第二写作信息,从而向用户展示该第二写作信息,可以便于用户进行思维发散与参考辅助写作。并且,无需用户逐个查看并筛选信息,可以提升用户的写作效率。

[0099] 本公开实施例的信息查询方法,通过根据待写作的目标事件查询各子图谱,以从各子图谱中确定第二目标子图谱,其中,第二目标子图谱中的第二节点的属性特征与目标事件匹配,第二节点的属性特征包括第一目标信息所属的事件;将第二目标子图谱,作为第二写作信息。由此,将与待写作的目标事件匹配的子图谱作为第二写作信息,从而向用户展示上述第二写作信息,可以便于用户进行思维发散与参考辅助写作,提升写作质量。并且,无需用户逐个查看并筛选信息,可以提升用户的写作效率。

[0100] 为了清楚说明本公开任一实施例中是如何确定各候选关键词对应的关联词的,本公开还提出一种信息查询方法。

[0101] 图5为本公开实施例四所提供的信息查询方法的流程示意图。

[0102] 如图5所示,该信息查询方法可以包括以下步骤:

[0103] 步骤501,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0104] 步骤502,对关键词集合中的任意两个关键词确定共同出现的次数,并根据上述次数,确定任意两个关键词的关联程度。

[0105] 在本公开实施例中,可以对关键词集合中的任意两个关键词(任意两个关键词即为任意一个关键词对),确定该任意两个关键词共同出现的次数,并根据上述次数,确定任意两个关键词的关联程度。其中,关联程度与次数成正向关系,即共同出现的次数越多,关联程度越高,反之,共同出现的次数越少,关联程度越低。

[0106] 作为一种示例,以第一候选信息为新闻资讯进行示例,可以统计任意两个关键词在各新闻资讯中共同出现的次数(或称为共现频次),根据共同出现的次数,确定任意两个关键词的关联程度。

[0107] 在本公开实施例的一种可能的实现方式中,为了简化计算,还可以对任意两个关键词共同出现的次数进行归一化处理,从而可以根据归一化后的次数,确定任意两个关键词的关联程度。其中,关联程度与归一化后的次数成正向关系,即归一化后的次数越多,关联程度越高,反之,归一化后的次数越少,关联程度越低。

[0108] 例如,可以确定所有的关键词对共同出现的次数的均值和均方差,归一化处理可以为:(次数-均值)/均方差。

[0109] 步骤503,针对关键词集合中的至少一个候选关键词,根据该候选关键词与关键词集合中除该候选关键词之外的其余关键词之间的关联程度,从其余关键词中确定与候选关键词关联的关联词。

[0110] 在本公开实施例中,针对关键词集合中的至少一个候选关键词,可以根据该候选关键词与关键词集合中除该候选关键词之外的其余关键词之间的关联程度,从其余关键词中确定与该候选关键词关联的关联词。

[0111] 作为一种示例,可以按照候选关键词与各其余关键词之间的关联程度的取值,由高至低地对各其余关键词进行排序,选取排序在前的第一设定个数的关键词,作为与该候选关键词关联的关联词。

[0112] 步骤504,确定与候选关键词关联的各第一目标信息,并确定与关联词关联的各第二目标信息。

[0113] 步骤505,根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0114] 步骤506,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0115] 步骤504至506的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0116] 本公开实施例的信息查询方法,通过对关键词集合中的任意两个关键词确定共同出现的次数,并根据上述次数,确定任意两个关键词的关联程度;针对关键词集合中的至少一个候选关键词,根据该候选关键词与关键词集合中除该候选关键词之外的其余关键词之间的关联程度,从其余关键词中确定与候选关键词关联的关联词。由此,可以实现各关键词共同出现的次数,有效确定与各关键词关联的关联词。

[0117] 为了清楚说明本公开任一实施例中是如何确定各候选关键词关联的第一目标信息的,本公开还提出一种信息查询方法。

[0118] 图6为本公开实施例五所提供的信息查询方法的流程示意图。

[0119] 如图6所示,该信息查询方法可以包括以下步骤:

[0120] 步骤601,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0121] 步骤602,针对关键词集合中的至少一个候选关键词,确定与该候选关键词关联的关联词。

[0122] 步骤601至602的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0123] 步骤603,确定与该候选关键词匹配的多个第二候选信息。

[0124] 在本公开实施例中,针对每个候选关键词,可以确定与该候选关键词匹配的多个第二候选信息。

[0125] 作为一种示例,以候选信息为新闻资讯进行示例,可以基于BM算法(Boyer-Moore,一种字符串搜索算法),比如BM25算法,在搜索服务器(比如ElasticSearch)中检索与该候选关键词相关的新闻资讯,作为与该候选关键词匹配的第二候选信息。

[0126] 步骤604,从多个第二候选信息中,确定与候选关键词关联的各第一目标信息;其中,各第一目标信息之间的相似度未高于相似度阈值,且各第一目标信息属于同一事件的概率未高于概率阈值。

[0127] 在本公开实施例中,可以从与候选关键词匹配的多个第二候选信息中,确定与候选关键词关联的各第一目标信息,其中,各第一目标信息之间的相似度未高于设定的相似

度阈值,且各第一目标信息属于同一事件的概率未高于设定的概率阈值。

[0128] 作为一种可能的实现方式,针对多个第二候选信息中的任意两个候选信息,可以确定该任意两个候选信息之间的第一相似度(比如文本相似度),并确定该任意两个候选信息属于同一事件的第一概率,比如可以基于文本分类算法,确定任意两个候选信息属于同一事件的第一概率。作为一种示例,可以基于经过训练的文本分类模型,识别任意两个候选信息属于同一事件的第一概率。

[0129] 例如,可以获取样本对(即信息对),其中,样本对可以包括属于同一事件的正样本对,和/或,不属于同一事件的负样本对,正样本对上标注有较高的标注概率,该标注概率用于指示正样本对中的两个信息属于同一事件的概率,负样本对上标注有较低的标注概率,该标注概率用于指示负样本对中的两个信息属于同一事件的概率。本公开中,可以将样本对输入文本分类模型,由文本分类模型输出样本对中两个信息属于同一事件的预测概率,从而可以根据预测概率和标注概率之间的差异,对文本分类模型进行训练,以使上述差异最小化。从而本公开中,可以采用经过训练后的文本分类模型,预测任意两个候选信息属于同一事件的第一概率。

[0130] 本公开实施例中,在第一相似度未高于设定的相似度阈值,且第一概率未高于设定的概率阈值的情况下,可以保留上述任意两个候选信息,而在第一相似度阈值高于设定的相似度阈值,和/或,第一概率高于设定的概率阈值的情况下,可以删除上述任意两个候选信息中的一个信息,从而可以将保留的各第二候选信息,作为与候选关键词关联的各第一目标信息。

[0131] 作为另一种可能的实现方式,还可以对候选关键词初始化一个信息集合,本公开中记为第一信息集合,并将多个第二候选信息中的任意一个信息,添加至该第一信息集合中,并计算各第二候选信息与第一信息集合中信息之间的相似度,以及属于同一事件的概率,如果某个第二候选信息和第一信息集合中信息之间的相似度未高于相似度阈值,且属于同一事件的概率未高于概率阈值,则将该第二候选信息添加至第一信息集合中。从而可以将第一信息集合中的各信息,作为与候选关键词关联的各第一目标信息。

[0132] 步骤605,确定与关联词关联的各第二目标信息。

[0133] 需要说明的是,本公开仅以步骤605在步骤604之后执行进行示例,但本公开并不限于此,步骤605还可以与步骤603至604并列执行,或者,步骤605还可以在步骤603之前执行,或者,步骤605可以在步骤603之后,且在步骤604之前执行,等等,即本公开对603-604,及605的执行时序不做限制,可以顺序执行,也可以并列执行。

[0134] 步骤606,根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0135] 步骤607,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0136] 步骤605至607的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0137] 本公开实施例的信息查询方法,通过确定与该候选关键词匹配的多个第二候选信息;从多个第二候选信息中,确定与候选关键词关联的各第一目标信息;其中,各第一目标信息之间的相似度未高于相似度阈值,且各第一目标信息属于同一事件的概率未高于概率

阈值。由此,可以实现对各第二候选信息进行去同处理,使得保留的各第一目标信息不仅与候选关键词关联,而且各第一目标信息之间的内容不相关,从而使得生成的子图谱简洁、不冗余。

[0138] 为了清楚说明本公开任一实施例中是如何确定与各候选关键词匹配的第二候选信息的,本公开还提出一种信息查询方法。

[0139] 图7为本公开实施例六所提供的信息查询方法的流程示意图。

[0140] 如图7所示,该信息查询方法可以包括以下步骤:

[0141] 步骤701,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0142] 步骤702,针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词。

[0143] 步骤701至702的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0144] 步骤703,搜索发布时间与当前时间处于设定范围内的第三候选信息。

[0145] 需要说明的是,某些信息(比如新闻资讯)是具有时效性的,为了便于用户参考辅助写作,提升写作质量,可以仅从全网搜索发布时间与当前时间处于设定范围内的信息,作为第三候选信息。

[0146] 步骤704,确定各第三候选信息与候选关键词之间的匹配度。

[0147] 在本公开实施例中,可以分别确定候选关键词与各第三候选信息之间的匹配度。比如,可以基于BM算法,确定各第三候选信息与候选关键词之间的匹配度。

[0148] 步骤705,根据对各第三候选信息监听到的用户交互行为,确定各第三候选信息的热度。

[0149] 在本公开实施例中,用户交互行为可以包括但不限于点击、浏览、评论、关注、收藏、分享、点赞等。

[0150] 在本公开实施例中,可以根据对第三候选信息监听到的用户交互行为,确定第三候选信息的热度。例如,可以根据各第三候选信息的浏览时长、点击量、评论量、关注量、收藏量、分享量、点赞量等信息,确定各第三候选信息的热度。其中,热度与第三候选信息的浏览时长、点击量、评论量、关注量、收藏量、分享量、点赞量等中的每一项成正向关系。

[0151] 步骤706,根据各第三候选信息的匹配度和/或热度,从各第三候选信息中确定与候选关键词匹配的多个第二候选信息。

[0152] 在本公开实施例的一种可能的实现方式中,可以根据各第三候选信息的匹配度,从各第三候选信息中确定与候选关键词匹配的多个第二候选信息。比如,可以选取匹配度较大的第二设定个数的第三候选信息,作为与候选关键词匹配的多个第二候选信息。

[0153] 在本公开实施例的另一种可能的实现方式中,可以根据各第三候选信息的热度,从各第三候选信息中确定与候选关键词匹配的多个第二候选信息。比如,可以选取热度较大的第三设定个数的第三候选信息,作为与候选关键词匹配的多个第二候选信息。

[0154] 在本公开实施例的另一种可能的实现方式中,还可以根据各第三候选信息的匹配度和热度,从各第三候选信息中确定与候选关键词匹配的多个第二候选信息。比如,可以选取热度较大,且匹配度较大的第四设定个数的第三候选信息,作为与候选关键词匹配的多个第二候选信息。

[0155] 其中,第一设定个数、第二设定个数、第三设定个数和第四设定个数可以根据需求设置,第一设定个数、第二设定个数、第三设定个数和第四设定个数的取值可以相同,或者也可以不同,本公开对此并不作限制。

[0156] 作为一种示例,针对每个第三候选信息,可以根据设定权重,对该第三候选信息的匹配度和热度进行加权求和,得到该第三候选信息的推荐度,从而可以选取推荐度较大的第四设定个数的第三候选信息,作为与候选关键词匹配的多个第二候选信息。

[0157] 步骤707,从多个第二候选信息中,确定与候选关键词关联的各第一目标信息;其中,各第一目标信息之间的相似度未高于相似度阈值,且各第一目标信息属于同一事件的概率未高于概率阈值。

[0158] 步骤708,确定与关联词关联的各第二目标信息。

[0159] 步骤709,根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0160] 步骤710,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0161] 步骤707至710的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0162] 本公开实施例的信息查询方法,通过搜索发布时间与当前时间处于设定范围内的第三候选信息;确定各第三候选信息与候选关键词之间的匹配度;根据对各第三候选信息监听到的用户交互行为,确定各第三候选信息的热度;根据各第三候选信息的匹配度和/或热度,从各第三候选信息中确定与候选关键词匹配的多个第二候选信息。由此,可以实现根据近期发布的热门信息,生成关键词对应的子图谱,可以使得用户及时了解近期的重大事件或重要事件,从而提升写作质量。

[0163] 为了清楚说明本公开任一实施例中是如何确定各关联词关联的第二目标信息的,本公开还提出一种信息查询方法。

[0164] 图8为本公开实施例七所提供的信息查询方法的流程示意图。

[0165] 如图8所示,该信息查询方法可以包括以下步骤:

[0166] 步骤801,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0167] 步骤802,针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词。

[0168] 步骤803,确定与候选关键词关联的各第一目标信息。

[0169] 步骤801至803的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0170] 步骤804,确定与关联词匹配的多个第四候选信息。

[0171] 在本公开实施例中,针对每个关联词,可以确定与该关联词匹配的多个第四候选信息。

[0172] 作为一种示例,以候选信息为新闻资讯进行示例,可以基于BM算法,比如BM25算法,在搜索服务器(比如ElasticSearch)中检索与该关联词相关的新闻资讯,作为与该关联词匹配的第四候选信息。

[0173] 在本公开实施例的一种可能的实现方式中,针对每个关联词,可以搜索发布时间

与当前时间处于设定范围内的第三候选信息,并确定各第三候选信息与该关联词之间的匹配度,并根据对各第三候选信息监听到的用户交互行为,确定各第三候选信息的热度,从而可以根据各第三候选信息的匹配度和/或热度,从各第三候选信息中确定与该关联词匹配的多个第四候选信息。

[0174] 步骤805,从多个第四候选信息中,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于相似度阈值,且各第二目标信息属于同一事件的概率未高于概率阈值。

[0175] 在本公开实施例中,可以从与关联词匹配的多个第四候选信息中,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于设定的相似度阈值,且各第二目标信息属于同一事件的概率未盖雨设定的概率阈值。

[0176] 步骤806,根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0177] 步骤807,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0178] 本公开实施例的信息查询方法,通过确定与关联词匹配的多个第四候选信息;从多个第四候选信息中,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于相似度阈值,且各第二目标信息属于同一事件的概率未高于概率阈值。由此,可以实现对各第四候选信息进行去同处理,使得保留的各第二目标信息不仅与关联词关联,而且各第二目标信息之间的内容不相关,从而使得生成的子图谱简洁、不冗余。

[0179] 为了清楚说明上述实施例中是如何确定与关联词关联的各第二目标信息的,本公开还提出一种信息查询方法。

[0180] 图9为本公开实施例八所提供的信息查询方法的流程示意图。

[0181] 如图9所示,该信息查询方法可以包括以下步骤:

[0182] 步骤901,对至少一个第一候选信息提取关键词,以得到关键词集合。

[0183] 步骤902,针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词。

[0184] 步骤903,确定与候选关键词关联的各第一目标信息。

[0185] 步骤904,确定与关联词匹配的多个第四候选信息。

[0186] 步骤901至904的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0187] 步骤905,确定各第一目标信息所属的第一类别,并从各第一类别中确定目标类别,其中,属于目标类别的第一目标信息个数大于其他类别。

[0188] 在本公开实施例中,可以根据文本分类算法,确定各第一目标信息所属的类别,本公开中记为第一类别。

[0189] 以信息为新闻资讯进行示例,信息所属的类别可以包括:社会、科技、娱乐等。假设第一目标信息的标题为“某明星在某地拍戏”,则该第一目标信息的类别可以为娱乐,假设第一目标信息的标题为“某个生活必需品涨价”,则该第一目标信息的类别可以为社会。

[0190] 在本公开实施例中,在确定各第一目标信息所属的第一类别后,可以从各第一类别中确定目标类别。其中,属于目标类别的第一目标信息个数大于其他类别。即,可以统计

每个第一类别所包含的第一目标信息的个数,根据各第一类别所包含的第一目标信息的个数,从各第一类别中确定目标类别,比如,属于目标类别的第一目标信息的个数可以位于前N个,N为正整数,比如,N可以为1、2、3等等。

[0191] 步骤906,确定多个第四候选信息所属的第二类别,并从多个第四候选信息中,确定第二类别与目标类别相同的第五候选信息。

[0192] 在本公开实施例中,可以基于文本分类算法,确定各第四候选信息所属的第二类别,并从各第四候选信息中,确定第二类别与目标类别相同的第五候选信息,即删除第二类别与目标类别不同的第四候选信息,将保留的各第四候选信息,作为第五候选信息。

[0193] 步骤907,根据各第五候选信息,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于相似度阈值,且各第二目标信息属于同一事件的概率未高于概率阈值。

[0194] 在本公开实施例中,可以根据各第五候选信息,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于设定的相似度阈值,且各第二目标信息属于同一事件的概率未高于设定的概率阈值。

[0195] 作为一种可能的实现方式,针对多个第五候选信息中的任意两个候选信息,可以确定该任意两个候选信息之间的第二相似度(比如文本相似度),并确定该任意两个候选信息属于同一事件的第二概率,比如可以基于文本分类算法,确定任意两个候选信息属于同一事件的第二概率。作为一种示例,可以基于经过训练的文本分类模型,识别任意两个候选信息属于同一事件的第二概率。

[0196] 本公开实施例中,在第二相似度未高于设定的相似度阈值,且第二概率未高于设定的概率阈值的情况下,可以保留上述任意两个候选信息,而在第二相似度阈值高于设定的相似度阈值,和/或,第二概率高于设定的概率阈值的情况下,可以删除上述任意两个候选信息中的一个信息,从而可以将保留的各第五候选信息,作为与关联词关联的各第二目标信息。

[0197] 作为另一种可能的实现方式,还可以对关联词初始化一个信息集合,本公开中记为第二信息集合,并将多个第五候选信息中的任意一个信息,添加至该第二信息集合中,并计算各第五候选信息与第二信息集合中信息之间的相似度,以及属于同一事件的概率,如果某个第五候选信息和第二信息集合中信息之间的相似度未高于相似度阈值,且属于同一事件的概率未高于概率阈值,则将该第五候选信息添加至第二信息集合中。从而可以将第二信息集合中的各信息,作为与关联词关联的各第二目标信息。

[0198] 步骤908,根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0199] 步骤909,根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0200] 步骤908至909的执行过程可以参见本公开任一实施例的执行过程,在此不做赘述。

[0201] 作为一种示例,以信息为新闻资讯进行示例性说明,可以通过以下几步,构建媒体行业的专题图谱:

[0202] 第一,对于每一篇新闻资讯,可以使用wordrank算法抽取该新闻资讯中的各关

关键词,并计算该新闻资讯所属的类别(即行业分类,包括社会、科技、娱乐等);

[0203] 第二,对于所有新闻资讯的关键词集合,统计每一对关键词在各篇新闻资讯中共同出现的次数(或称为共现频次),并对共现频次进行归一化,作为关键词对的关联程度;

[0204] 第三,对于关键词集合中要计算关系图谱的候选关键词,后续称为中心词,可以根据关联程度,确定与该中心词关联程度最高的6个关键词,作为与该中心词关联的关联词;

[0205] 第四,对于中心词和各关联词中的任意一个关键词,采用BM25算法,在ElasticSearch中检索一周内与该关键词相关的新闻资讯,并选取相关的新闻资讯中热度处于top 50的资讯,作为候选新闻资讯,其中,各候选新闻资讯可以按照热度降序排列;

[0206] 第五,针对中心词和各关联词中的每个关键词,初始化一个信息集合,将该关键词对应的每个候选新闻资讯,与信息集合中的资讯计算文本相似度,以及属于同一事件的概率(或称为事件同属分数,可以基于事件同属判别技术,判断两个资讯的标题是否属于同一个事件),如果文本相似度和事件同属分数都低于阈值,则将该候选新闻资讯插入到信息集合中;

[0207] 第六,统计中心词对应的信息集合中各候选新闻资讯中大多数资讯所属的类别(后续称为行业分类),并选取top3的类别作为公共行业分类;对于关联词对应的信息集合中各候选新闻资讯,如果行业分类和公共行业分类无交集,则从信息集合中删除该候选新闻资讯,取信息集合中剩余的资讯作为与关联词关联的资讯;

[0208] 第七,根据中心词和关联词之间的关联关系、中心词与对应的信息集合中各候选新闻资讯之间的关联关系,以及关联词与对应的信息集合中各候选新闻资讯之间的关联关系,生成该中心词对应的子图谱。

[0209] 其中,每个中心词对应的子图谱可以每隔几小时更新一次。在计算完每个中心词的子图谱之后,可以将各中心词的子图谱保存到数据库中,供检索接口使用。

[0210] 相比传统的人工构建专题图谱的方式,本公开中通过自动构建新闻专题图谱,来发现实体、事件之间的关联性,不仅有助于帮助内容创作者寻找创作灵感,提升创作效率,而且还可以节省人力成本和时间成本,提升图谱中关联关系的覆盖率。即通过分析新闻资讯的关键词、事件、实体等信息,分析关键词之间的隐式关系,构建起节点之间的树形关系,得到一个个的小专题,小专题的集合构成整个媒体行业的专题图谱,帮助创作者发现内容之间的关联性从而进行深度写作。

[0211] 一个典型的场景是,媒体创作者想要写作关于“XX运动会”的深度文章,需要搜集跟“XX运动会”关系比较密切的信息。通过检索媒体行业专题图谱,可以获取与“XX运动会”强相关的关键词,以及与关键词有关的事件,可以将专题图谱提供的信息,用于写作前的思维发散和写作时候的内容信息。

[0212] 本公开实施例的信息查询方法,通过确定各第一目标信息所属的第一类别,并从各第一类别中确定目标类别,其中,属于目标类别的第一目标信息个数大于其他类别;确定多个第四候选信息所属的第二类别,并从多个第四候选信息中,确定第二类别与目标类别相同的第五候选信息;根据各第五候选信息,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于相似度阈值,且各第二目标信息属于同一事件的概率未高于概率阈值。由此,根据与候选关键词关联的信息所属的类别,对关联词关联的信息进行筛选,可以使得生成的子图谱中各节点均与中心节点(即候选关键词)强相关,从而提升

候选关键词对应的子图谱生成的准确性和可靠性。

[0213] 与上述图1至图9实施例提供的信息查询方法相对应,本公开还提供一种信息查询装置,由于本公开实施例提供的信息查询装置与上述图1至图9实施例提供的信息查询方法相对应,因此在信息查询方法的实施方式也适用于本公开实施例提供的信息查询装置,在本公开实施例中不再详细描述。

[0214] 图10为本公开实施例九所提供的信息查询装置的结构示意图。

[0215] 如图10所示,该信息查询装置1000可以包括:提取模块1001、第一确定模块1002、第二确定模块1003、第三确定模块1004、生成模块1005以及处理模块1006。

[0216] 其中,提取模块1001,用于对至少一个第一候选信息提取关键词,以得到关键词集合。

[0217] 第一确定模块1002,用于针对关键词集合中的至少一个候选关键词,确定与候选关键词关联的关联词。

[0218] 第二确定模块1003,用于确定与候选关键词关联的各第一目标信息。

[0219] 第三确定模块1004,用于确定与关联词关联的各第二目标信息。

[0220] 生成模块1005,用于根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱。

[0221] 处理模块1006,用于根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。

[0222] 在本公开实施例的一种可能的实现方式中,生成模块1005,具体用于:将候选关键词作为对应的子图谱中的中心节点;将关联词作为与中心节点关联的第一节点;根据各第一目标信息的属性特征,确定与中心节点关联的第二节点;根据各第二目标信息的属性特征,确定与第一节点关联的第三节点。

[0223] 在本公开实施例的一种可能的实现方式中,属性特征,用于指示第一目标信息或第二目标信息所属的事件,处理模块1006,还用于:根据待写作的目标事件查询各子图谱,以从各子图谱中确定第二目标子图谱,其中,第二目标子图谱中的第二节点的属性特征与目标事件匹配,第二节点的属性特征包括第一目标信息所属的事件;将第二目标子图谱,作为第二写作信息。

[0224] 在本公开实施例的一种可能的实现方式中,第一确定模块1002,具体用于:对关键词集合中的任意两个关键词确定共同出现的次数,并根据次数,确定任意两个关键词的关联程度;根据候选关键词与关键词集合中除候选关键词之外的其余关键词之间的关联程度,从其余关键词中确定与候选关键词关联的关联词。

[0225] 在本公开实施例的一种可能的实现方式中,第二确定模块1003,具体用于:确定与候选关键词匹配的多个第二候选信息;从多个第二候选信息中,确定与候选关键词关联的各第一目标信息;其中,各第一目标信息之间的相似度未高于相似度阈值,且各第一目标信息属于同一事件的概率未高于概率阈值。

[0226] 在本公开实施例的一种可能的实现方式中,第二确定模块1003,具体用于:搜索发布时间与当前时间处于设定范围内的第三候选信息;确定各第三候选信息与候选关键词之间的匹配度;根据对各第三候选信息监听到的用户交互行为,确定各第三候选信息的热度;根据各第三候选信息的匹配度和/或热度,从各第三候选信息中确定与候选关键词匹配的

多个第二候选信息。

[0227] 在本公开实施例的一种可能的实现方式中,第三确定模块1004,具体用于:确定与关联词匹配的多个第四候选信息;从多个第四候选信息中,确定与关联词关联的各第二目标信息,其中,各第二目标信息之间的相似度未高于相似度阈值,且各第二目标信息属于同一事件的概率未高于概率阈值。

[0228] 在本公开实施例的一种可能的实现方式中,第三确定模块1004,具体用于:确定各第一目标信息所属的第一类别,并从各第一类别中确定目标类别,其中,属于目标类别的第一目标信息个数大于其他类别;确定多个第四候选信息所属的第二类别,并从多个第四候选信息中,确定第二类别与目标类别相同的第五候选信息;根据各第五候选信息,确定与关联词关联的各第二目标信息。

[0229] 本公开实施例的信息查询装置,通过对至少一个第一候选信息提取关键词,以得到关键词集合,针对关键词集合中的至少一个候选关键词,确定与该候选关键词关联的关联词,并确定与候选关键词关联的各第一目标信息,及与关联词关联的各第二目标信息;根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱;根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。由此,通过根据各关键词之间的关联关系,关键词与信息之间的关联关系,自动生成各关键词对应的子图谱,而无需人工整理的方式得到,可以提升图谱的生成效率,降低人工成本。并且,将与待写作的目标关键词对应的第一目标子图谱,作为第一写作信息,从而向用户展示该第一写作信息,可以便于用户进行思维发散与参考辅助写作。此外,无需用户逐个查看并筛选信息,可以提升用户的写作效率。

[0230] 为了实现上述实施例,本公开还提供一种电子设备,该电子设备可以包括至少一个处理器;以及与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行本公开上述任一实施例提出的信息查询方法。

[0231] 为了实现上述实施例,本公开还提供一种存储有计算机指令的非瞬时计算机可读存储介质,其中,计算机指令用于使计算机执行本公开上述任一实施例提出的信息查询方法。

[0232] 为了实现上述实施例,本公开还提供一种计算机程序产品,该计算机程序产品包括计算机程序,计算机程序在被处理器执行时实现本公开上述任一实施例提出的信息查询方法。

[0233] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0234] 图11示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。其中,电子设备可以包括上述实施例中的服务端、客户端。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0235] 如图11所示,电子设备1100包括计算单元1101,其可以根据存储在ROM(Read-Only Memory,只读存储器)1102中的计算机程序或者从存储单元1108加载到RAM(Random Access Memory,随机访问/存取存储器)1103中的计算机程序,来执行各种适当的动作和处理。在RAM 1103中,还可存储电子设备1100操作所需的各种程序和数据。计算单元1101、ROM 1102以及RAM 1103通过总线1104彼此相连。I/O(Input/Output,输入/输出)接口1105也连接至总线1104。

[0236] 电子设备1100中的多个部件连接至I/O接口1105,包括:输入单元1106,例如键盘、鼠标等;输出单元1107,例如各种类型的显示器、扬声器等;存储单元1108,例如磁盘、光盘等;以及通信单元1109,例如网卡、调制解调器、无线通信收发机等。通信单元1109允许电子设备1100通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0237] 计算单元1101可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元1101的一些示例包括但不限于CPU(Central Processing Unit,中央处理单元)、GPU(Graphic Processing Units,图形处理单元)、各种专用的AI(Artificial Intelligence,人工智能)计算芯片、各种运行机器学习模型算法的计算单元、DSP(Digital Signal Processor,数字信号处理器)、以及任何适当的处理器、控制器、微控制器等。计算单元1101执行上文所描述的各个方法和处理,例如上述信息查询方法。例如,在一些实施例中,上述信息查询方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元1108。在一些实施例中,计算机程序的部分或者全部可以经由ROM 1102和/或通信单元1109而被载入和/或安装到电子设备1100上。当计算机程序加载到RAM 1103并由计算单元1101执行时,可以执行上文描述的信息查询方法的一个或多个步骤。备选地,在其他实施例中,计算单元1101可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行上述信息查询方法。

[0238] 本文中以上描述的系统和技术和各种实施方式可以在数字电子电路系统、集成电路系统、FPGA(Field Programmable Gate Array,现场可编程门阵列)、ASIC(Application-Specific Integrated Circuit,专用集成电路)、ASSP(Application Specific Standard Product,专用标准产品)、SOC(System On Chip,芯片上系统的系统)、CPLD(Complex Programmable Logic Device,复杂可编程逻辑设备)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0239] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0240] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可

读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、RAM、ROM、EPROM (Electrically Programmable Read-Only-Memory,可擦除可编程只读存储器)或快闪存储器、光纤、CD-ROM (Compact Disc Read-Only Memory,便捷式紧凑盘只读存储器)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0241] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT (Cathode-Ray Tube,阴极射线管)或者LCD (Liquid Crystal Display,液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0242] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:LAN (Local Area Network,局域网)、WAN (Wide Area Network,广域网)、互联网和区块链网络。

[0243] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务 (Virtual Private Server,虚拟专用服务器)中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0244] 其中,需要说明的是,人工智能是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科,既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术;人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

[0245] 根据本公开实施例的技术方案,通过对至少一个第一候选信息提取关键词,以得到关键词集合,针对关键词集合中的至少一个候选关键词,确定与该候选关键词关联的关联词,并确定与候选关键词关联的各第一目标信息,及与关联词关联的各第二目标信息;根据候选关键词、关联词、各第一目标信息和各第二目标信息,生成候选关键词对应的子图谱;根据待写作的目标关键词查询各子图谱,以将与目标关键词对应的第一目标子图谱,作为第一写作信息。由此,通过根据各关键词之间的关联关系,关键词与信息之间的关联关系,自动生成各关键词对应的子图谱,而无需人工整理的方式得到,可以提升图谱的生成效率,降低人工成本。并且,将与待写作的目标关键词对应的第一目标子图谱,作为第一写作

信息,从而向用户展示该第一写作信息,可以便于用户进行思维发散与参考辅助写作。此外,无需用户逐个查看并筛选信息,可以提升用户的写作效率。

[0246] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开提出的技术方案所期望的结果,本文在此不进行限制。

[0247] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

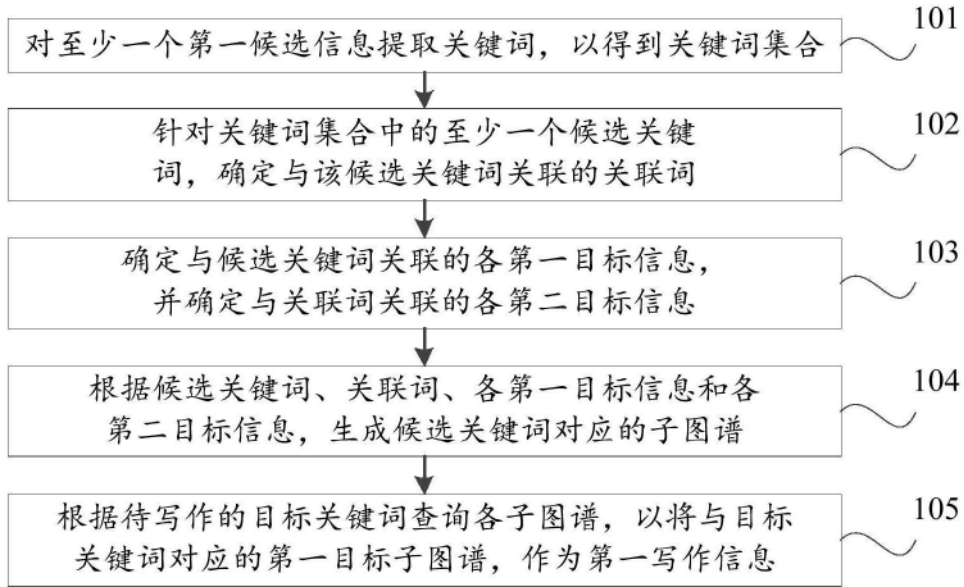


图1

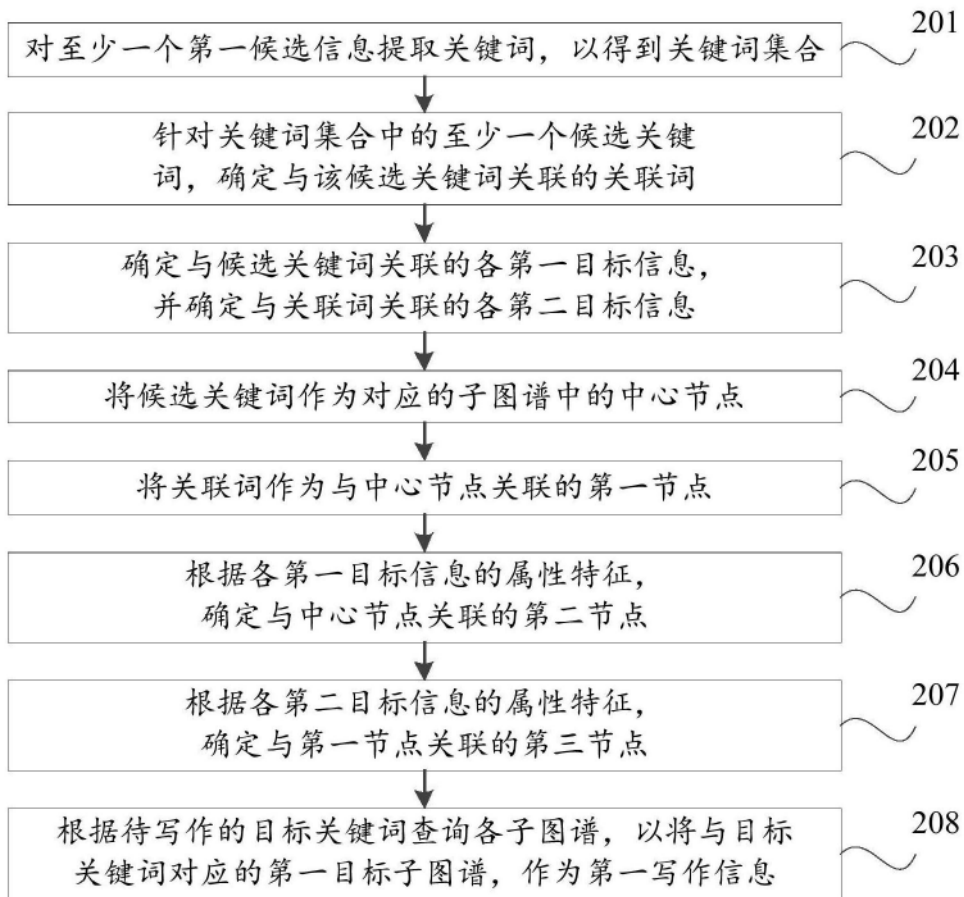


图2

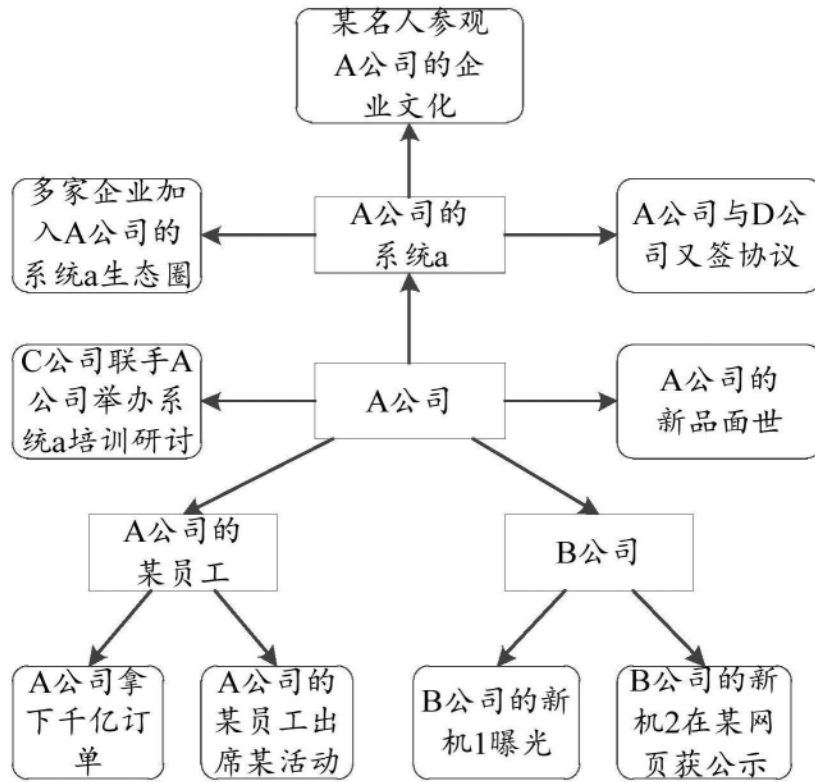


图3

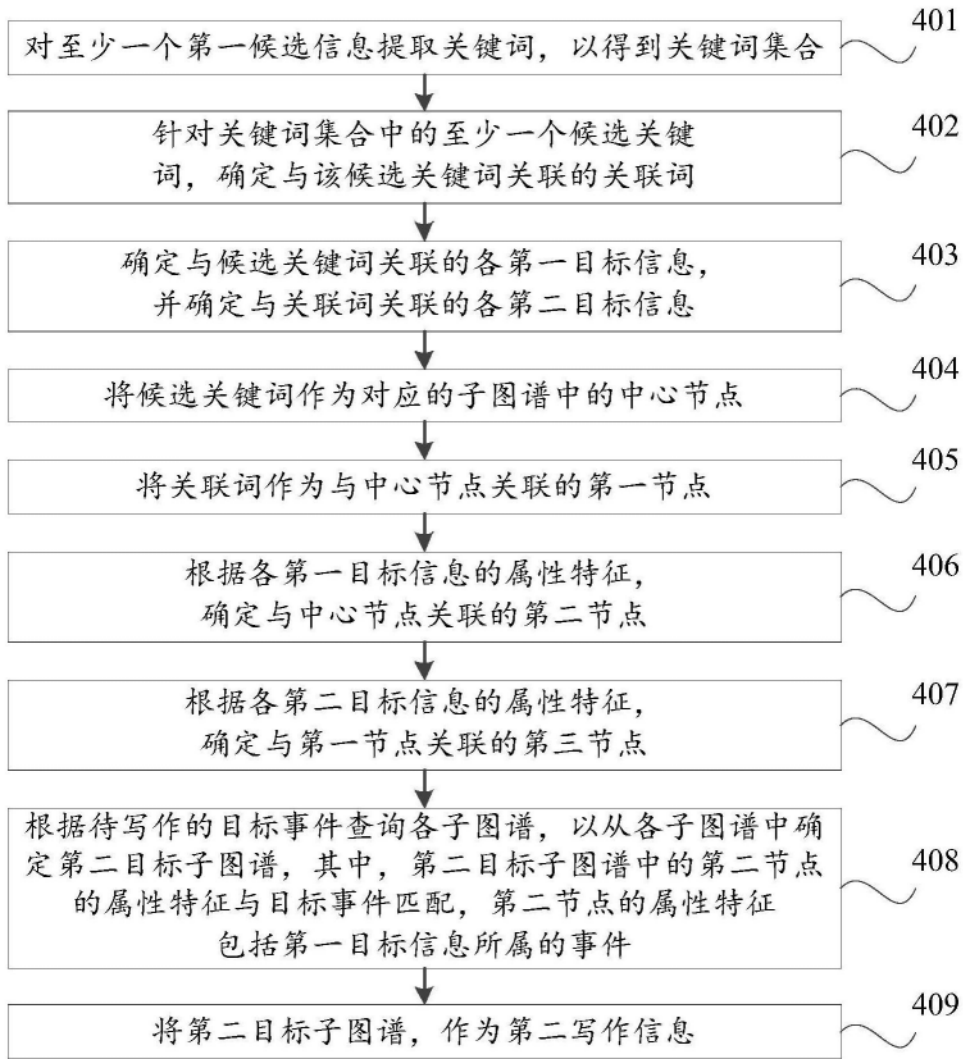


图4

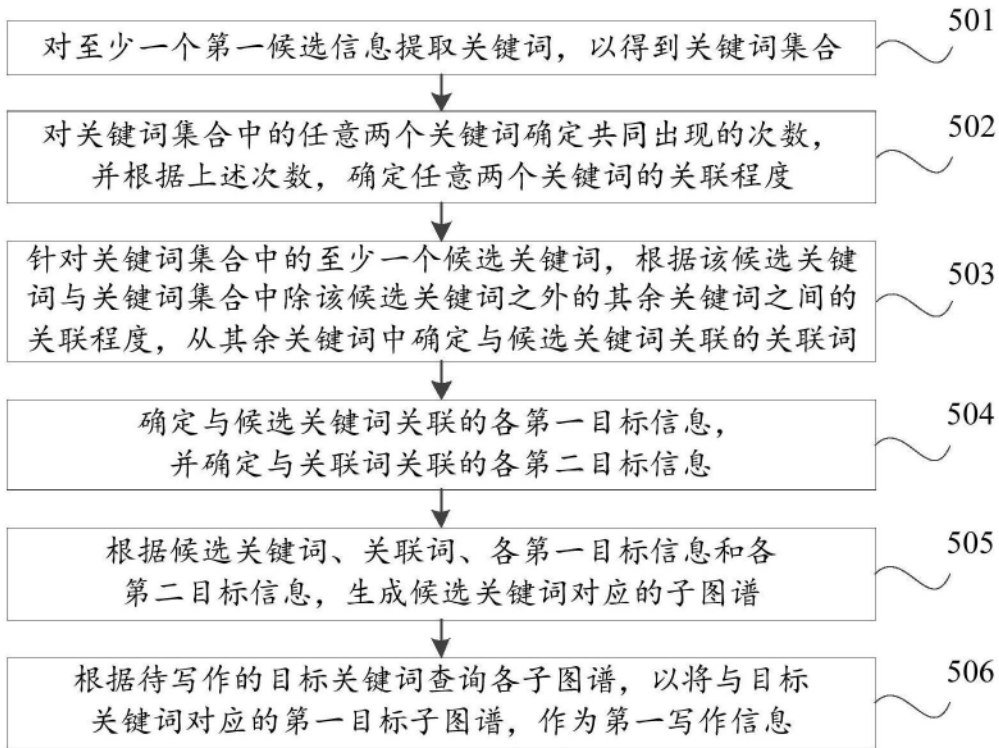


图5

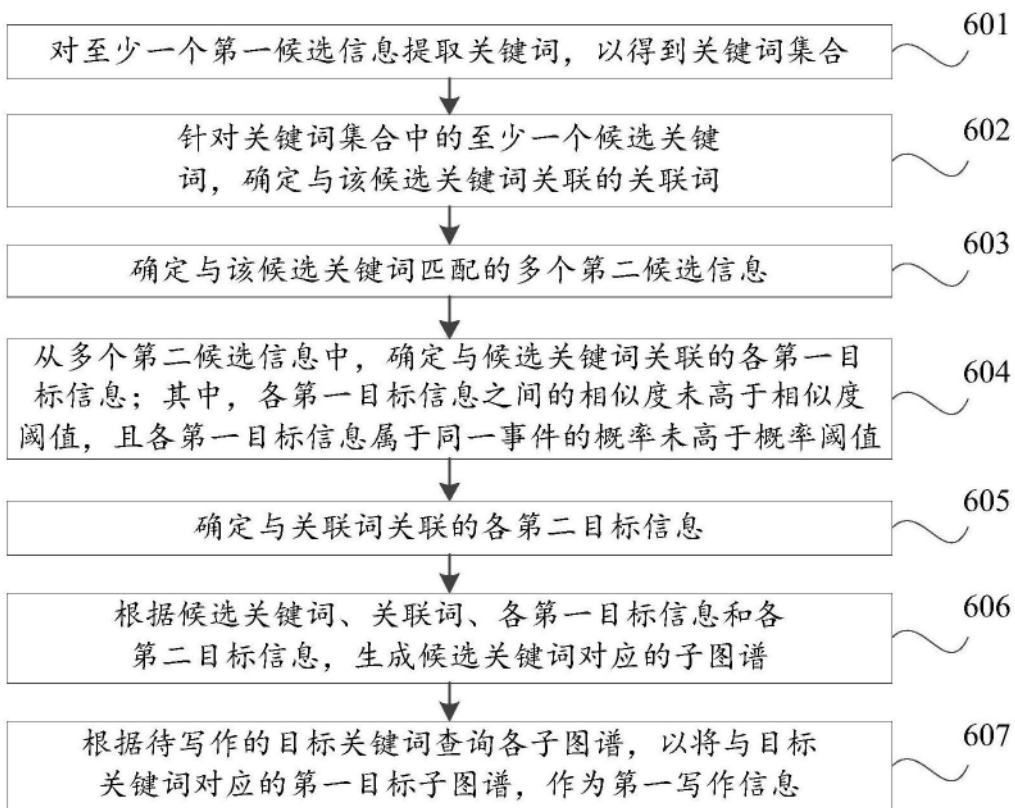


图6

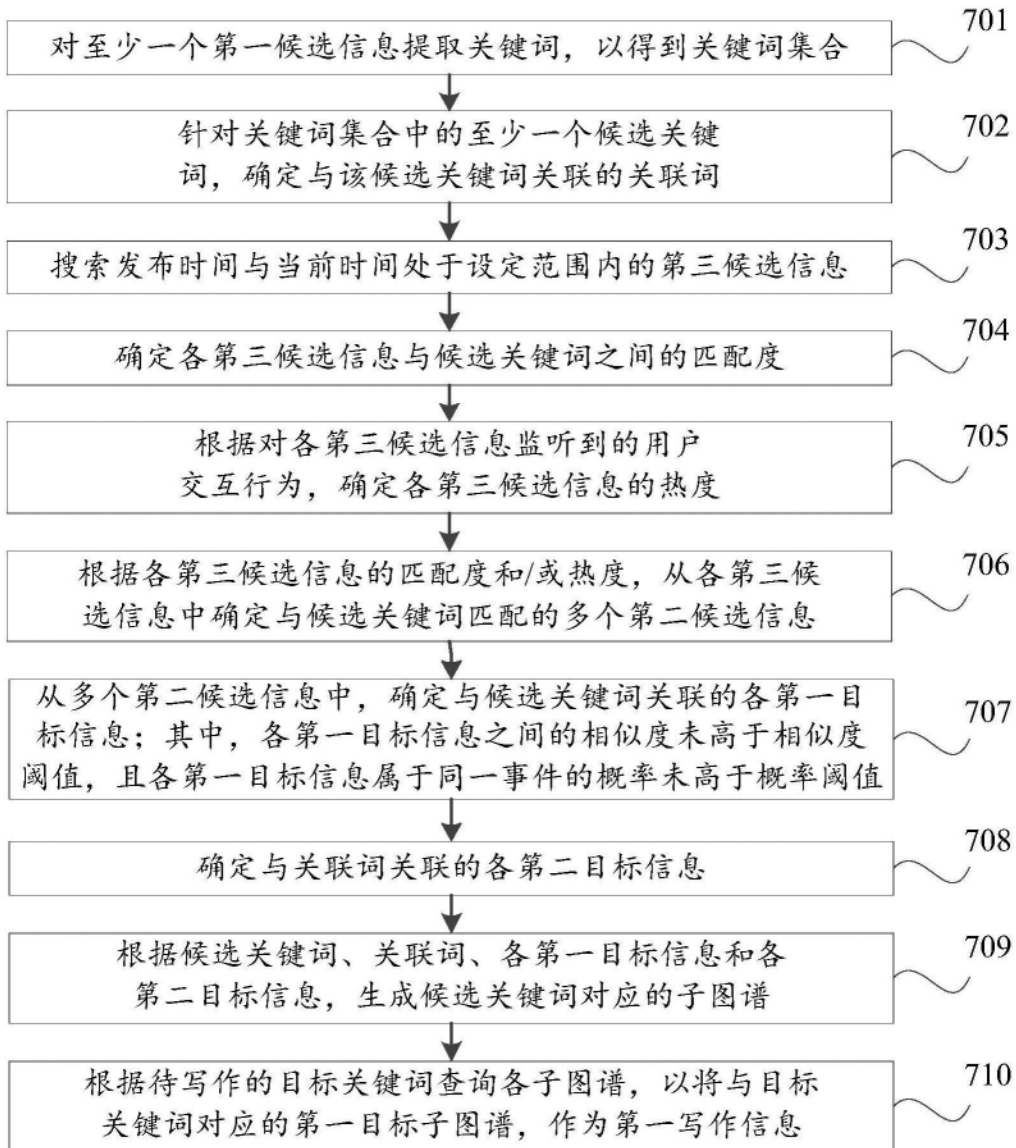


图7

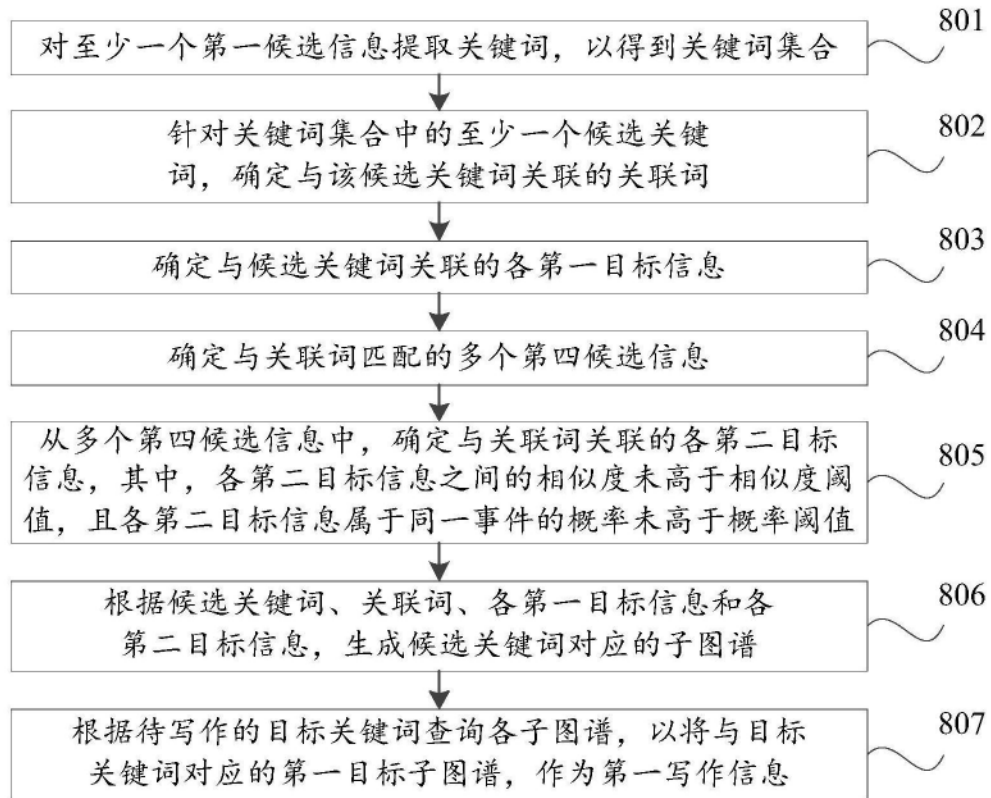


图8

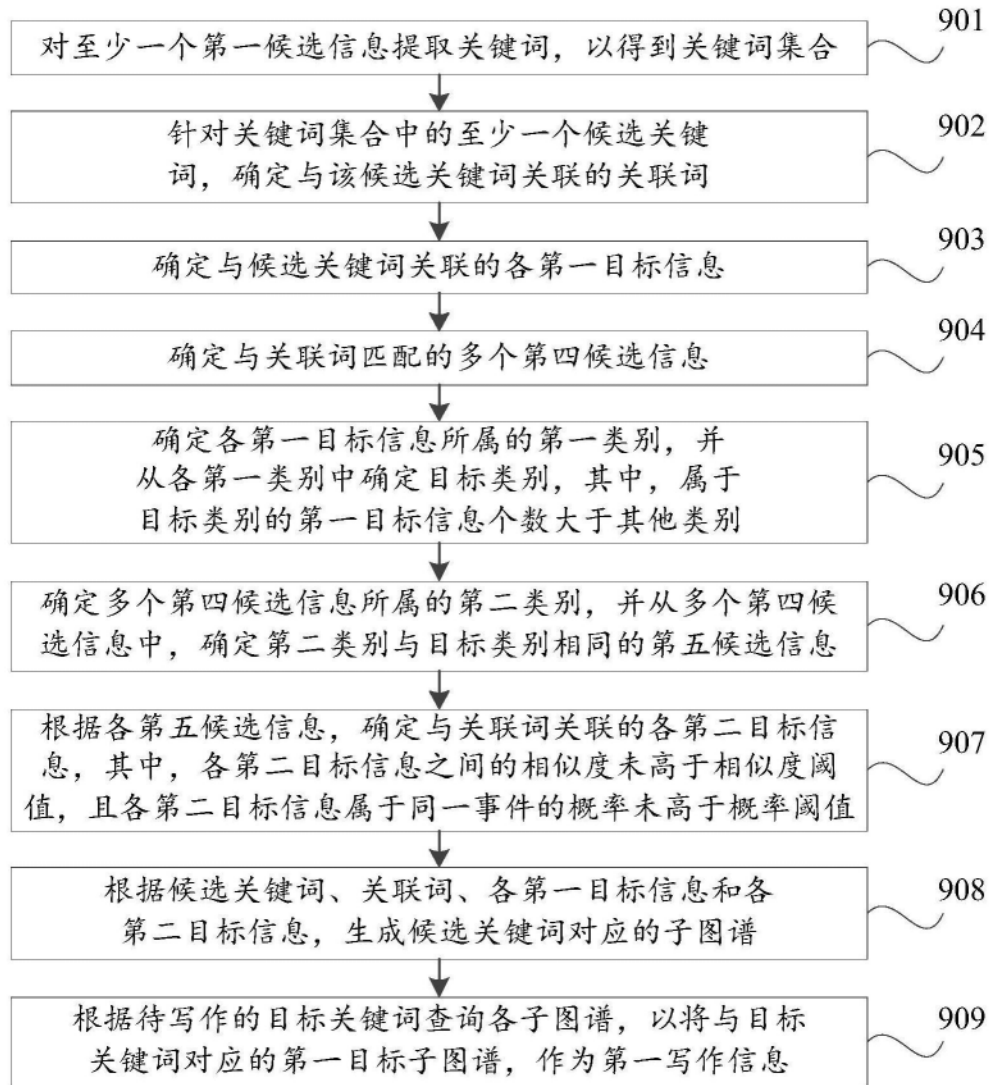


图9

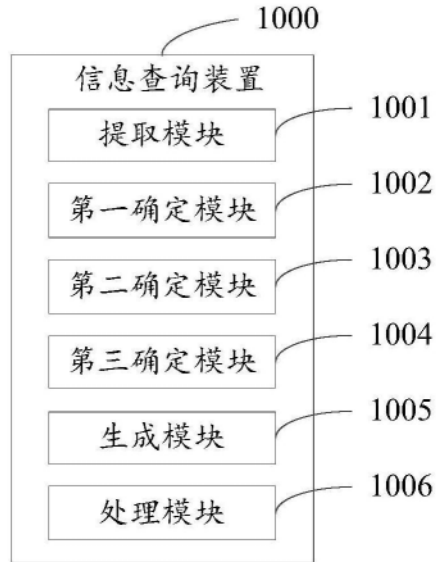


图10

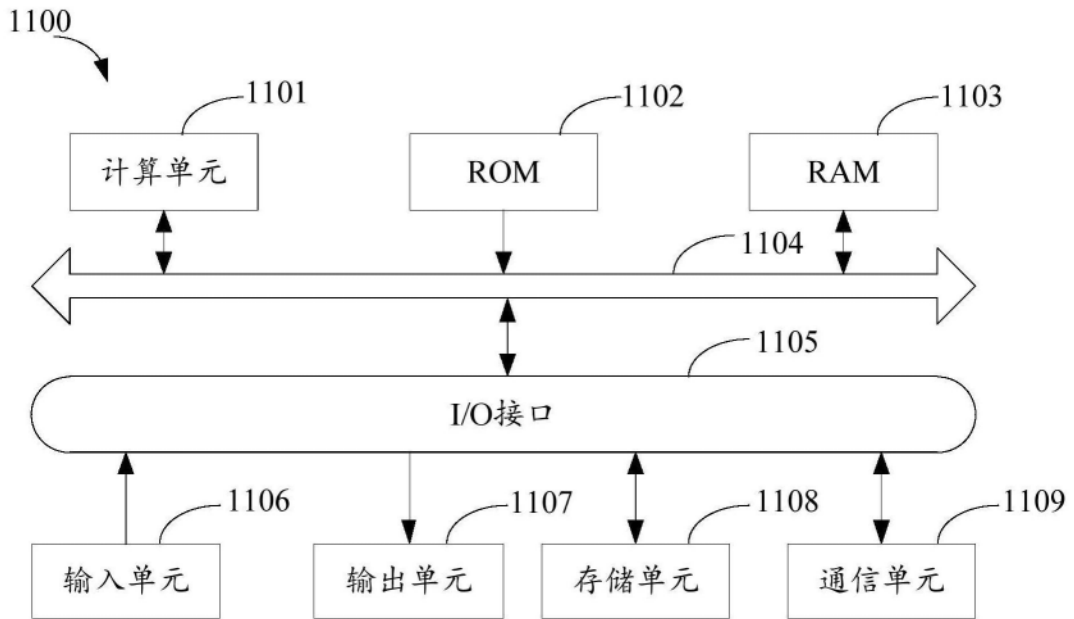


图11