



(12) 发明专利

(10) 授权公告号 CN 111753198 B

(45) 授权公告日 2024. 01. 12

(21) 申请号 202010577208.5

G06N 3/045 (2023.01)

(22) 申请日 2020.06.22

G06N 3/0442 (2023.01)

(65) 同一申请的已公布的文献号

G06N 20/00 (2019.01)

申请公布号 CN 111753198 A

G06F 40/216 (2020.01)

G06F 40/289 (2020.01)

(43) 申请公布日 2020.10.09

G06F 40/30 (2020.01)

(73) 专利权人 北京百度网讯科技有限公司

(56) 对比文件

地址 100085 北京市海淀区上地十街10号

WO 2020009027 A1, 2020.01.09

百度大厦2层

JP 2019074843 A, 2019.05.16

(72) 发明人 王海威 王杰 刘准 朱勇 田浩

US 2020012733 A1, 2020.01.09

和为

CN 110188168 A, 2019.08.30

JP 2019053520 A, 2019.04.04

(74) 专利代理机构 中科专利商标代理有限责任

WO 2014127500 A1, 2014.08.28

公司 11021

US 2019102430 A1, 2019.04.04

专利代理师 吕朝蕙

夏向阳; 肖文. 社会科学领域国际专利信息

(51) Int. Cl.

研究的知识图谱分析. 浙江大学学报(人文社会科学版). 2019, (06), 全文.

G06F 16/9535 (2019.01)

G06F 16/35 (2019.01)

G06F 16/36 (2019.01)

G06N 3/0464 (2023.01)

审查员 王志超

权利要求书3页 说明书17页 附图10页

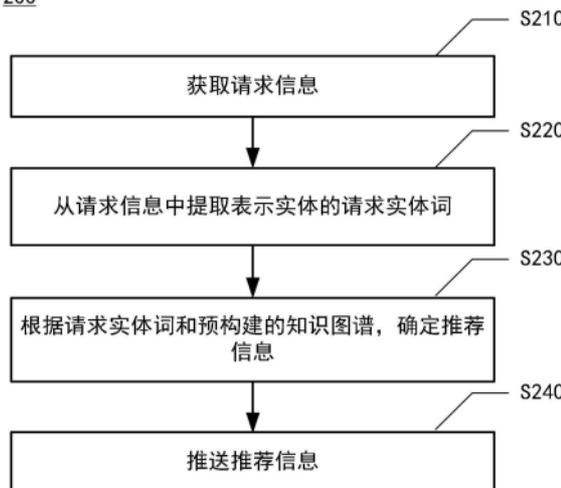
(54) 发明名称

信息推荐方法和装置、以及电子设备和可读存储介质

(57) 摘要

本公开提供了一种信息推荐方法, 涉及知识图谱领域。该方法包括: 获取请求信息; 从请求信息中提取表示实体的请求实体词; 根据请求实体词和预构建的知识图谱, 确定推荐信息; 以及推送推荐信息, 其中, 知识图谱是基于文本构建的, 且知识图谱指示了对文本的来源加以表示的第一词。本公开还提供了一种信息推荐装置、一种电子设备以及一种计算机可读存储介质。

200



1. 一种信息推荐方法,包括:
 - 获取请求信息;
 - 从所述请求信息中提取表示实体的请求实体词;
 - 从所述请求信息中提取关键信息;
 - 确定预设关键信息结构中与所述提取的关键信息相关联的信息节点,作为目标信息节点,所述关键信息结构包括多个信息节点,所述多个信息节点中的每个信息节点用于指示一个关键信息,所述关键信息结构中的每个信息节点对应有一个存储空间,用于存储被归类至该每个信息节点的文本;
 - 根据所述请求实体词、预构建的知识图谱以及所述目标信息节点,确定推荐信息;以及推送所述推荐信息,
 - 其中,所述知识图谱是基于文本构建的,且所述知识图谱指示了对所述文本的来源加以表示的第一词,
 - 其中,所述确定推荐信息包括:
 - 获取归类至所述目标信息节点的文本,作为所述推荐信息。
2. 根据权利要求1所述的方法,还包括基于文本构建知识图谱,包括,
 - 从所述文本中提取表示实体的多个实体词以及所述多个实体词彼此之间的关联关系,所述多个实体词包括所述第一词;以及
 - 根据所提取的所述多个实体词和所述多个实体词彼此之间的关联关系,构建知识图谱。
3. 根据权利要求2所述的方法,其中,所述构建知识图谱包括:针对所述多个实体词中的每个实体词,
 - 针对所述每个实体词构建节点;
 - 根据所述多个实体词彼此之间的关联关系,确定所述多个实体词中与所述每个实体词具有关联关系的关联词;以及
 - 连接针对所述每个实体词的节点与针对所述关联词的节点,以形成与针对所述每个实体词的节点相关联的边。
4. 根据权利要求3所述的方法,其中,所述构建知识图谱还包括:
 - 根据所述每个实体词与所述关联词之间的关联关系,确定所述每个实体词与所述关联词之间的关联强度;以及
 - 根据所述关联强度,向连接针对所述每个实体词的节点与针对所述关联词的节点的边分配权重,
 - 其中,所述关联强度的大小与所述权重的大小成正比。
5. 根据权利要求3所述的方法,其中,所述确定推荐信息包括:
 - 确定所述知识图谱中针对所述请求实体词的节点,作为目标节点;
 - 根据与所述目标节点相关联的边,确定与所述目标节点连接的至少一个节点;以及
 - 根据所述至少一个节点所针对的实体词,确定所述推荐信息。
6. 根据权利要求2所述的方法,还包括:
 - 响应于获取到新文本,从所述新文本中提取表示实体的多个新实体词以及所述多个新实体词彼此之间的关联关系;以及

在所述知识图谱不能指示所述多个新实体词中至少一个新实体词的情况下,根据所述多个新实体词及所述多个新实体词彼此之间的关联关系,更新所述知识图谱。

7. 根据权利要求2所述的方法,其中,所述多个实体词还包括第二词,所述第二词用于表示所述文本的关键信息;

所述方法还包括:将所述文本归类为所述关键信息结构包括的信息节点,包括,根据所述第二词,确定所述文本与所述关键信息结构中每个信息节点之间的匹配关系;以及

将所述文本归类为与所述文本相匹配的信息节点。

8. 根据权利要求7所述的方法,其中,所述第二词的数量为多个;所述确定所述文本与所述关键信息结构中每个信息节点的匹配关系包括:

确定所述第二词是否包括目标词,所述目标词与所述每个信息节点指示的关键信息相同;

在所述第二词中包括所述目标词的情况下,确定所述文本与所述每个信息节点之间的匹配关系为相匹配。

9. 根据权利要求8所述的方法,其中,所述确定所述文本与所述关键信息结构中每个信息节点之间的匹配关系还包括:

在所述第二词中不包括目标词的情况下,从所述知识图谱中获取所述第二词的关联词;以及

在所述第二词的关联词包括所述目标词的情况下,确定所述文本与所述每个信息节点之间的匹配关系为相匹配。

10. 根据权利要求7所述的方法,其中,所述确定所述文本与所述关键信息结构中每个信息节点之间的匹配关系包括:

确定所述第二词与所述每个信息节点指示的关键信息之间的相似度;以及

在所述相似度大于相似度阈值的情况下,确定所述文本与所述每个信息节点之间的匹配关系为相匹配。

11. 根据权利要求7所述的方法,还包括:

响应于获取到新文本,从所述新文本中提取所述新文本的新关键信息;以及

在所述关键信息结构中不包括指示所述新关键信息的信息节点的情况下,根据所述新关键信息,更新所述关键信息结构。

12. 一种信息推荐装置,包括:

获取模块,用于获取请求信息;

信息提取模块,用于从所述请求信息中提取表示实体的请求实体词;以及从所述请求信息中提取关键信息;

节点确定模块,用于确定预设关键信息结构中与所述提取的关键信息相关联的信息节点,作为目标信息节点,所述关键信息结构包括多个信息节点,所述多个信息节点中的每个信息节点用于指示一个关键信息,所述关键信息结构中的每个信息节点对应有一个存储空间,用于存储被归类至该每个信息节点的文本

信息确定模块,用于根据所述请求实体词、预构建的知识图谱以及所述目标信息节点,确定推荐信息;以及

信息推送模块,用于推送所述推荐信息,

其中,所述知识图谱是基于文本构建的,且所述知识图谱指示了对所述文本的来源加以表示的第一词,

其中,所述信息确定模块还用于:获取归类至所述目标信息节点的文本,作为所述推荐信息。

13.一种计算机系统,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

其中,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述一个或多个处理器执行权利要求1~11中任一项所述的方法。

14.一种计算机可读存储介质,其上存储有可执行指令,该指令被处理器执行时使处理器执行权利要求1~11中任一项所述的方法。

信息推荐方法和装置、以及电子设备和可读存储介质

技术领域

[0001] 本公开涉及知识图谱领域,更具体地,涉及一种信息推荐方法和装置、以及电子设备和可读存储介质。

背景技术

[0002] 随着互联网技术的发展,为了提高办公效率且便于企业管理,各种协助管理的人工智能平台应运而生。

[0003] 在实现本公开构思的过程中,发明人发现相关技术中至少存在以下技术问题:相关技术中的人工智能平台,多以信息体系构建、信息分类及信息分享为核心。但在响应于检索语句进行信息推荐时,往往仅能根据检索语句向用户推荐包括检索语句中关键词的信息。因此,相关技术中的人工智能平台,存在缺少根据检索语句对用户的意图进行深层次挖掘和理解,从而不能向用户推荐更丰富的信息的技术问题。

发明内容

[0004] 有鉴于此,本公开提供了一种能够深层次的挖掘和理解用户需求,向用户推荐丰富信息的信息推荐方法和装置、以及电子设备和可读存储介质。

[0005] 本公开的一个方面提供了一种信息推荐方法,该方法包括:获取请求信息;从请求信息中提取表示实体的请求实体词;根据请求实体词和预构建的知识图谱,确定推荐信息;以及推送推荐信息,其中,知识图谱是基于文本构建的,且知识图谱指示了对文本的来源加以表示的第一词。

[0006] 可选地,上述信息推荐方法还包括基于文本构建知识图谱,包括:从文本中提取表示实体的多个实体词以及多个实体词彼此之间的关联关系,多个实体词包括所述第一词;以及根据所提取的多个实体词和多个实体词彼此之间的关联关系,构建知识图谱。

[0007] 可选地,上述构建知识图谱包括:针对多个实体词中的每个实体词,针对每个实体词构建节点;根据多个实体词彼此之间的关联关系,确定多个实体词中与每个实体词具有关联关系的关联词;以及连接针对每个实体词的节点与针对关联词的节点,以形成与针对每个实体词的节点相关联的边。

[0008] 可选地,上述构建知识图谱还包括:根据每个实体词与关联词之间的关联关系,确定每个实体词与关联词之间的关联强度;以及根据关联强度,向连接针对每个实体词的节点与针对关联词的节点的边分配权重,其中,关联强度的大小与权重的大小成正比。

[0009] 可选地,上述确定推荐信息包括:确定知识图谱中针对请求实体词的节点,作为目标节点;根据与目标节点相关联的边,确定与目标节点连接的至少一个节点;以及根据至少一个节点所针对的实体词,确定推荐信息。

[0010] 可选地,上述信息推荐方法还包括:响应于获取到新文本,从新文本中提取表示实体的多个新实体词及多个新实体词彼此之间的关联关系;以及在知识图谱不能指示多个新实体词中至少一个新实体词的情况下,根据多个新实体词及多个新实体词彼此之间的关联

关系,更新知识图谱。

[0011] 可选地,上述信息推荐方法还包括:从请求信息中提取关键信息;以及确定预设关键信息结构中与所述提取的关键信息相关联的信息节点,作为目标信息节点,关键信息结构包括多个信息节点,多个信息节点中的每个信息节点用于指示一个关键信息,其中,确定推荐信息包括:根据请求实体词、预构建的知识图谱以及目标知识节点,确定推荐信息。

[0012] 可选地,多个实体词还包括第二词,该第二词用于表示文本的关键信息;上述信息推荐方法还包括:将文本归类为关键信息结构包括的信息节点,包括:根据第二词,确定文本与关键信息结构中每个信息节点之间的匹配关系;以及将文本归类为与文本相匹配的信息节点。

[0013] 可选地,第二词的数量为多个;确定文本与关键信息结构中每个信息节点的匹配关系包括:确定第二词是否包括目标词,目标词与每个信息节点指示的关键信息相同;在第二词中包括目标词的情况下,确定文本与每个知识节点之间的匹配关系为相匹配。

[0014] 可选地,确定文本与关键信息结构中每个信息节点之间的匹配关系还包括:在第二词中不包括目标词的情况下,从知识图谱中获取第二词的关联词;以及在第二词的关联词包括目标词的情况下,确定文本与每个信息节点之间的匹配关系为相匹配。

[0015] 可选地,确定文本与关键信息结构中每个信息节点之间的匹配关系包括:确定第二词与每个信息节点指示的关键信息之间的相似度;以及在相似度大于相似度阈值的情况下,确定文本与每个信息节点之间的匹配关系为相匹配。

[0016] 可选地,上述信息推荐方法还包括:响应于获取到新文本,从新文本中提取新文本的新关键信息;以及在关键信息结构中不包括指示新关键信息的信息节点的情况下,根据新关键信息,更新关键信息结构。

[0017] 可选地,确定推荐信息包括:获取归类至目标信息节点的文本,作为推荐信息。

[0018] 本公开的另一方面提供了一种信息推荐装置,该装置包括:获取模块,用于获取请求信息;信息提取模块,用于从请求信息中提取表示实体的请求实体词;信息确定模块,用于根据请求实体词和预构建的知识图谱,确定推荐信息;以及信息推送模块,用于推送推荐信息,其中,知识图谱是基于文本构建的,且知识图谱指示了对文本的来源加以表示的第一词。

[0019] 本公开的另一方面提供了一种计算机系统,该计算机系统包括:一个或多个处理器;以及存储装置,用于存储一个或多个程序。其中,当一个或多个程序被一个或多个处理器执行时,使得一个或多个处理器执行如上所述的信息推荐方法。

[0020] 本公开的另一方面提供了一种计算机可读存储介质,其上存储有可执行指令,该指令被处理器执行时使处理器执行如上所述的信息推荐方法。

[0021] 本公开的另一方面提供了一种计算机程序,该计算机程序包括计算机可执行指令,该指令在被执行时用于实现如上所述的信息推荐方法。

[0022] 本公开的另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上述方法。

[0023] 根据本公开的实施例,可以至少部分地避免相关信息推荐方法存在的推荐信息单一、无法深入企业团队和项目管理的缺陷。并因此通过能够指示文本来源的知识图谱来进行信息推荐,可以使得推荐结果能够深入企业团队和项目管理中。从而有效提高企业知识

应用的效率,提高员工办公效率,提高人工智能平台对知识的智能化管理。

附图说明

[0024] 通过以下参照附图对本公开实施例的描述,本公开的上述以及其他目的、特征和优点将更为清楚,在附图中:

[0025] 图1示意性示出了根据本公开实施例的信息推荐方法和装置、以及电子设备和可读存储介质的应用场景;

[0026] 图2示意性示出了根据本公开实施例的信息推荐方法的流程图;

[0027] 图3A示意性示出了根据本公开实施例的基于文本构建知识图谱的流程图;

[0028] 图3B示意性示出了根据本公开实施例的用于构建知识图谱的文本的获取流程示意图;

[0029] 图3C示意性示出了根据本公开另一实施例的基于文本构建知识图谱的流程图;

[0030] 图4A示意性示出了根据本公开实施例的确定推荐信息的流程图;

[0031] 图4B示意性示出了根据本公开实施例的根据知识图谱确定推荐信息的原理示意图;

[0032] 图5示意性示出了根据本公开实施例的更新知识图谱的流程图;

[0033] 图6A示意性示出了根据本公开另一实施例的信息推荐方法的流程图;

[0034] 图6B示意性示出了根据本公开实施例的预设关键信息结构的示意图;

[0035] 图7示意性示出了根据本公开实施例的将文本归类为关键信息结构包括的信息节点的流程图;

[0036] 图8示意性示出了根据本公开实施例的确定文本与关键信息结构中每个信息节点之间的匹配关系的流程图;

[0037] 图9示意性示出了根据本公开另一实施例的确定文本与关键信息结构中每个信息节点之间的匹配关系的流程图;

[0038] 图10示意性示出了根据本公开实施例的用于实现信息推荐方法的系统架构图;

[0039] 图11示意性示出了根据本公开实施例的信息推荐装置的结构框图;以及

[0040] 图12示意性示出了根据本公开实施例的适于执行信息推荐方法的计算机系统的方框图。

具体实施方式

[0041] 以下,将参照附图来描述本公开的实施例。但是应该理解,这些描述只是示例性的,而并非要限制本公开的范围。在下面的详细描述中,为便于解释,阐述了许多具体的细节以提供对本公开实施例的全面理解。然而,明显地,一个或多个实施例在没有这些具体细节的情况下也可以被实施。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本公开的概念。

[0042] 在此使用的术语仅仅是为了描述具体实施例,而并非意在限制本公开。在此使用的术语“包括”、“包含”等表明了所述特征、步骤、操作和/或部件的存在,但是并不排除存在或添加一个或多个其他特征、步骤、操作或部件。

[0043] 在此使用的所有术语(包括技术和科学术语)具有本领域技术人员通常所理解的

含义,除非另外定义。应注意,这里使用的术语应解释为具有与本说明书的上下文相一致的含义,而不应以理想化或过于刻板的方式来解释。

[0044] 在使用类似于“A、B和C等中至少一个”这样的表述的情况下,一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释(例如,“具有A、B和C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等)。

[0045] 本公开实施例提供了一种用于提高推荐的信息的丰富性,提高用户体验的信息推荐方法和装置。该信息推荐方法可以先获取客户端接收的请求信息。然后响应于获取到请求信息,从请求信息中提取表示实体的请求实体词。随后根据请求实体词和预先构建的知识图谱,来确定推荐信息。最后将推荐信息推送给客户端,以向用户展示该推荐信息。其中知识图谱是基于文本构建的,且知识图谱指示了对文本的来源加以表示的第一词。通过该方法,可以根据指示文本来源的知识图谱,向用户推荐其所需求的知识的来源。便于用户快速了解到研究相关知识的团队、专家等信息,提升知识开发和应用的效率。

[0046] 图1示意性示出了根据本公开实施例的信息推荐方法和装置、以及电子设备和可读存储介质的应用场景。需要注意的是,图1所示仅为可以应用本公开实施例的系统架构的示例,以帮助本领域技术人员理解本公开的技术内容,但并不意味着本公开实施例不可以用于其他设备、系统、或场景。

[0047] 如图1所示,该实施例的应用场景100可以包括服务器101、网络102和终端设备103、104、105。网络102用以在服务器101和终端设备103、104、105之间提供通信链路的介质。网络102可以包括各种连接类型,例如有线或无线通信链路等等。

[0048] 终端设备103、104、105例如可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。该终端设备103、104、105可以响应于用户操作生成请求信息,并通过与服务器101的交互获取响应于该请求信息的推荐信息,将该推荐信息展示给用户。

[0049] 示例性地,终端设备103、104、105可以安装有各种客户端应用,例如文本编辑类应用、音频编辑类应用、视频编辑类应用、即时通信类应用、智能办公平台等(仅为示例)。

[0050] 服务器101例如可以根据获取的信息维护有知识管理系统。其中,获取的信息例如可以包括从终端设备103、104、105安装的多个客户端应用获取的文本信息、视频信息和/或音频信息等。该知识管理系统例如可以根据获取的信息构建并维护知识图谱,以将多个信息彼此之间的关联关系形成可视化结构,便于进行信息挖掘和推荐。

[0051] 示例性地,该服务器101还可以响应于获取到终端设备103、104、105生成的请求信息,根据其构建并维护的知识图谱确定向用户推荐的推荐信息,并将该推荐信息推送给终端设备103、104、105,供用户浏览。

[0052] 示例性地,该服务器101例如还可以为企业设置的知识管理服务器,其对企业内部员工上传的信息和线上办公时产生的信息进行管理和维护,知识图谱不仅可以根据获取的信息来构建,还可以同时根据获取的信息的来源来构建。其中,获取的信息的来源可以包括:上传信息的用户信息、生成信息的作者和/或团队等信息。在基于知识图谱确定推荐信息时,可以挖掘企业内员工要素与知识间的关联关系,以此提升企业知识生产和应用的效率。其中,服务器管理和维护的信息例如可以包括信息106,该信息106可以包括自Wiki系统

中获取的信息、进行线上会议时产生的音频和/或文本信息、通过线上办公平台传输的文件、通过线上办公平台生成的聊天信息等。

[0053] 需要说明的是,本公开实施例所提供的信息推荐方法一般可以由服务器101执行。相应地,本公开实施例所提供的信息推荐装置一般可以设置于服务器101中。

[0054] 应该理解,图1中的终端设备、网络和服务器的数目和类型仅仅是示意性的。根据实现需要,可以具有任意数目和类型的终端设备、网络和服务器。

[0055] 以下将基于图1的应用场景,结合图2~图9对本公开实施例的信息推荐方法进行详细描述。

[0056] 图2示意性示出了根据本公开实施例的信息推荐方法的流程图。

[0057] 如图2所示,该实施例的信息推荐方法200可以包括操作S210~操作S240。

[0058] 在操作S210,获取请求信息。

[0059] 根据本公开的实施例,请求信息例如可以是终端设备响应于用户操作生成的。终端设备在生成请求信息后,将该请求信息发送给服务器,以向服务器请求推荐信息。

[0060] 示例性地,用户操作例如可以为用户通过输入设备输入检索语句的操作。或者,用户操作可以为用户向终端设备发出语音指令的操作等。终端设备生成的请求信息为能够指示检索语句或语音指令、且服务器能够识别的信息。例如,该请求信息例如可以为一条query(查询)语句。

[0061] 在操作S220,从请求信息中提取表示实体的请求实体词。

[0062] 根据本公开的实施例,该操作S220例如可以先对请求信息进行分词处理,得到指示请求信息的词序列。然后对该词序列中的词进行实体识别,提取得到表示实体的词,得到请求实体词。示例性地,该请求实体词例如可以是指请求信息中的信息元素,如人名、组织/机构名、物体名称、项目名称等。

[0063] 示例性地,该操作S220例如可以通过预先训练的机器学习模型来进行请求实体词的提取。机器学习模型例如可以包括LSTM(Long Short-Term Memory,长短期记忆网络)等。

[0064] 示例性地,该操作S220例如可以通过将词序列中的每个词与实体词库中的词进行比对,从词序列包括的多个词中筛选得到属于实体词的词,该筛选得到的词即为实体的请求实体词。

[0065] 在操作S230,根据请求实体词和预构建的知识图谱,确定推荐信息。

[0066] 根据本公开的实施例,知识图谱是基于大量的文本构建得到的。例如,知识图谱可以是根据从文本中提取到的实体词和实体词表示的实体间的关联关系构成的。知识图谱中包括多个节点和连接节点的多条边。每个节点针对一个实体词,即每个节点可以表示一个实体。每条边用于指示连接的两个节点分别表示的两个实体之间具有关联关系。

[0067] 示例性地,为了响应于获取请求,向用户提供更丰富的信息,知识图谱包括的多个节点中,例如可以包括有用于指示对文本的来源加以表示的第一词。第一词例如可以包括文本的作者名称、生成文本的团队名称、组织名称等。在请求实体词中包括“Java”时,根据知识图谱确定的推荐信息例如可以包括企业内精通Java语言的专家、团队等。

[0068] 根据本公开的实施例,该操作S230例如可以先确定知识图谱中表示请求实体词的节点。然后根据确定的节点在知识图谱中的位置,确定与请求实体词表示的实体具有关联关系的实体。最后根据该具有关联关系的实体来生成推荐信息。

[0069] 根据本公开的实施例,服务器例如还可以维护有文本库,该文本库中包括终端设备上传的多个文本。该操作S230例如也可以先根据请求实体词,从文本库中获取与请求实体词匹配的文本。然后再确定知识图谱中与请求实体词具有关联关系的实体。最后根据从文本库中获取的文本及具有关联关系的实体词生成推荐信息。

[0070] 示例性地,知识图谱例如可以通过图3A描述的流程构建得到,操作S230可以采用图4A描述的流程来实现,在此不再详述。

[0071] 在操作S240,推送推荐信息。该操作可以是将生成的推荐信息推送给终端设备,以使得终端设备将推荐信息展示给用户。

[0072] 综上可知,本公开实施例在进行信息推荐时,通过基于指示有文本来源的知识图谱来确定推荐信息,可以向用户推荐针对请求信息的来源信息,例如领域专家或团队等。因此,可以使得该信息推荐能够深入企业团队和项目管理中,从而能够为用户推荐更全面更准确的信息。

[0073] 为了便于根据知识图谱进行信息推荐,该实施例的信息推荐方法还可以包括在操作S210之前执行的以下操作:基于文本构建知识图谱。

[0074] 图3A示意性示出了根据本公开实施例的基于文本构建知识图谱的流程图。图3B示意性示出了根据本公开实施例的用于构建知识图谱的文本的获取流程示意图。

[0075] 如图3A所示,基于文本构建知识图谱的操作S350例如可以包括操作S351~操作S352。

[0076] 在操作S351,从文本中提取表示实体的多个实体词及多个实体词彼此之间的关联关系。

[0077] 根据本公开的实施例,为了使得构建的知识图谱涵盖更多的知识,用于提取实体词和实体词之间的关联关系的文本例如可以包括从分散化的多个渠道获取的多个文本。该操作S351用于对多个文本中的每个文本进行实体词和关联关系的提取。在从多个文本中的每个文本提取到实体词和关联关系后,可以对多个文本中提取到的实体词和关联关系进行统计,得到用于构建知识图谱的多个实体词和多个实体词彼此之间的关联关系。

[0078] 示例性地,如图3B所示,多个文本例如可以指根据在线信息301、离线信息302、隐形信息303、缺失信息304生成的在线文本305。

[0079] 示例性地,在知识图谱是针对企业的知识图谱时,在线信息301例如可以通过以下方式得到:对企业内知识源进行梳理排查,收集得到企业内的在线信息。知识源例如包括:线上的新闻通知、线上文档、线上会议纪要、线上VCR、线上项目介绍、线上聊天内容等。

[0080] 示例性地,离线信息302可以包括存储于员工办公设备中的未上传的信息。该离线信息302可以通过督促员工上传来实现信息的线上化处理(操作S311),将离线信息转换为在线信息。

[0081] 示例性地,隐形信息303例如可以包括员工未文档化的工作经验等。该隐形信息例如可以通过激励的方式激励员工进行文档化处理(操作S312),即鼓励员工记录工作经验并将记录的工作经验电子化。该文档化后的隐形信息可以通过线上化处理(操作S311),转换为在线信息。

[0082] 示例性地,缺失信息304例如可以包括企业内部没有而需要从外部引入的信息。该缺失信息304例如可以通过外部抓取、定向引入等方式从企业外部导入至企业内。例如,可

以通过调用公开的知识平台(论坛、论文网站、专利网站等)提供的接口,将缺失的信息导入企业内部的服务器中。或者,该缺失信息也可以通过运营生产(操作S313)的方式来鼓励员工创作得到。创作得到的信息首先需要文档化处理(操作S312),再通过线上化处理(操作S311),转换为在线信息。

[0083] 通过将离线信息302、隐形信息303、缺失信息304均转换为在线信息后,可以将通过对知识源进行梳理排查得到的在线信息301和转换得到的在线信息均进行文本化,得到多个在线文本305。

[0084] 根据本公开的实施例,为了在使得构建的知识图谱涵盖更多的知识的基础上,提高知识图谱的构建效率,该实施例还可以在得到多个在线文本305后,对多个在线文本305进行筛选,得到价值性高的在线文本。

[0085] 示例性地,如图3B所示,在得到在线文本305后,可以对多个在线文本进行知识性筛选(操作S314)和内容质量筛选(操作S315)。知识性筛选例如可以是筛选得到知识性较强的文本。内容质量筛选例如可以是筛选得到内容质量较高的文本。内容质量筛选可以是对知识性筛选得到的文本进行进一步筛选。或者,知识性筛选可以是对内容质量筛选得到的文本进行进一步筛选。

[0086] 知识性筛选例如可以通过对在线文本进行识别和理解后进行。例如,可以先对多个在线文本进行文本识别和文本理解,确定多个在线文本中每个在线文本描述的内容的类别。在文本属于能够为员工的能力提升提供帮助的文本类别(例如专业知识类别等)时,文本的知识性强。在文本属于不能够为员工的能力提供帮助的文本类别(例如八卦类别等)时,文本的知识性弱。

[0087] 内容质量筛选可以用于对文本中包括的词语进行分类后再进行。例如,若文本中包括的敏感词、禁用词等词语较多时,文本的内容质量较低。若文本中不包括敏感词、禁用词时,文本的质量较高。

[0088] 根据本公开的实施例,操作S351例如可以先采用类似于操作S220的方法从每个文本提取得到表示实体的实体词。然后再从文本中提取两个实体词分别表示的两个实体之间的语义关系,将该语义关系作为两个实体词之间的关联关系。

[0089] 根据本公开的实施例,在构建的知识图谱指示了对文本的来源加以表示的第一词时,从文本中提取的实体词应包括表示文本来源的第一词。在文本中不包括文本来源信息时,除了需要从文本中提取实体词外,例如还可以从文本的属性信息中提取实体词。文本的属性信息例如可以包括有作者信息、创建日期信息等。

[0090] 示例性地,操作S351可以通过深度学习的方法提取到每个文本中的多个实体词以多个实体词彼此之间的关联关系。深度学习的方法例如可以通过采用卷积神经网络模型来实现。

[0091] 示例性地,操作S351可以通过预定义的模板来提取文本中的实体词。预定义的模板规定了实体词抽取的规则等。

[0092] 在操作S352,根据所提取的所述多个实体词和多个实体词彼此之间的关联关系,构建知识图谱。

[0093] 根据本公开的实施例,该操作S352例如可以先针对操作S351中提取的多个实体词中的每个实体词构建节点。然后再根据多个实体词彼此之间的关联关系,确定多个实体词

中与每个实体词具有关联关系的关联词。最后连接针对每个实体词的节点与针对关联词的节点,形成与针对每个实体词的节点相关联的边。

[0094] 示例性地,对于实体词“Java”,若多个实体词中与“Java”具有关联关系的关联词包括“JavaWeb”、“Android”、“王**”,则该操作S352可以包括:连接针对“Java”的实体词的节点与针对“JavaWeb”的实体词的节点,连接针对“Java”的实体词的节点与针对“Android”的实体词的节点,连接针对“Java”的实体词的节点与针对“王**”的实体词的节点,形成与针对实体词“Java”的节点相关联的三条边。其中,“王**”可以是精通Java语言的专家的姓名,或者精通Android开发的专家的姓名等。

[0095] 根据本公开的实施例,为了便于在进行信息推荐时提高信息推荐的准确性,该实施例在构建知识图谱时,还可以为连接两个节点的每条边分配权重。

[0096] 图3C示意性示出了根据本公开另一实施例的基于文本构建知识图谱的流程图。

[0097] 如图3C所示,该实施例的基于文本构建知识图谱的操作S350除了操作S351~操作S352外,还可以包括操作S353~操作S354。该操作S353~操作S354在操作S352之后执行。

[0098] 在操作S353,根据每个实体词与关联词之间的关联关系,确定每个实体词与关联词之间的关联强度。

[0099] 在操作S354,根据关联强度,向连接针对每个实体词的节点与针对关联词的节点的边分配权重。

[0100] 根据本公开的实施例,操作S353例如可以根据每个实体词表示的实体与关联词表示的实体之间在多个维度中的关联关系,来确定每个词与关键词之间的关联强度。若每个实体词表示的实体与关联词表示的实体在较多维度中具有关联关系,则实体词与关联词之间的关联强度较大。若每个实体词表示的实体与关联词表示的实体在较少维度中具有关联关系,则实体词与关联词之间的关联强度较小。其中,多个维度例如可以包括地域维度、时间维度、种类维度、颜色维度、用户体验维度等。

[0101] 示例性地,对于实体词“苹果”,若多个实体词中与“苹果”具有关联关系的关联词包括“梨”、“橘子”。苹果和梨均属于水果、两者均产自北方,苹果和梨两者之间在两个维度上具有关联关系。苹果和橘子均属于水果,但苹果产自北方,橘子产自南方。苹果和橘子两者之间在一个维度上具有关联关系。因此,实体词“苹果”与实体词“梨”之间的关联强度大于实体词“苹果”与实体词“橘子”之间的关联强度。

[0102] 根据本公开的实施例,该操作S354可以向关联强度大的两个实体词所针对的两个节点之间的边分配较大的权重,而向关联强度小的两个实体词所针对的两个节点之间的边分配较小的权重。例如向针对实体词“苹果”的节点与针对实体词“梨”的节点之间的边分配的权重可以为第一权重,向针对实体词“苹果”的节点与针对实体词“橘子”的节点之间的边分配的权重可以为第二权重,此处的第一权重大于第二权重。

[0103] 示例性地,向两个节点之间的边分配的权重的大小可以与该两个节点所针对的两个实体词之间的关联强度的大小成正比。

[0104] 本公开实施例通过为连接两个节点的边分配权重,可以在根据请求实体词和预构建的知识图谱来确定推荐信息时,通过根据一端与针对请求实体词的节点连接的边的权重,优先根据权重大的边的另一端所连接的节点所针对的关联词来确定推荐信息。使得确定的推荐信息与请求信息匹配度高,并因此提高用户体验。

[0105] 图4A示意性示出了根据本公开实施例的确定推荐信息的流程图。图4B示意性示出了根据本公开实施例的根据知识图谱确定推荐信息的原理示意图。

[0106] 根据本公开的实施例,如图4A所示,在构建了知识图谱后,确定推荐信息的操作S230可以包括操作S431~操作S433。

[0107] 在操作S431,确定知识图谱中针对请求实体词的节点,作为目标节点。

[0108] 根据本公开的实施例,可以通过将请求实体词与知识图谱中各节点所针对的实体词进行比对,确定所针对的实体词与请求实体词一致的节点为目标节点。

[0109] 根据本公开的实施例,考虑到存在多个实体词用于表示同一实体的情况,还可以先计算请求实体词与知识图谱中各节点所针对的实体词之间的第一相似度,确定与请求实体词的第一相似度大于第一相似度阈值的实体词对应的节点为目标节点。其中,第一相似度例如可以采用余弦相似度、杰卡德相似系数等任意一种参数形式来体现。第一相似度阈值可以根据实际需求进行设定,例如,在一实施例中,该第一相似度阈值可以为0.8。

[0110] 在操作S432,根据与目标节点相关联的边,确定与目标节点连接的至少一个节点。

[0111] 示例性地,如图4B所示,通过操作S431确定的目标节点例如可以包括节点401和节点402。根据与节点401和节点402相关联的边,可以确定得到与目标节点连接的至少一个节点(图4B中右上角的阴影覆盖的节点),包括节点403、节点404、节点405和节点406。需要说明的是,如4B中示出的知识图谱的边是有指向性的,因此,在确定与目标节点连接的至少一个节点时,具体可以是与目标节点连接的边所指向的节点为至少一个节点。根据实际应用场景,本公开实施例知识图谱中的边可以是没有指向性的。

[0112] 在操作S433,根据至少一个节点所针对的实体词,确定推荐信息。

[0113] 根据本公开的实施例,可以根据操作S432确定的至少一个节点所针对的实体词直接生成推荐信息。例如,根据至少一个节点所针对的实体词及实体词所表示的实体的类别生成推荐信息。示例性地,对于请求实体词“Java”,若通过操作S432确定的至少一个节点所针对的实体词包括有“Android”和“王**”,“Android”属于操作系统类别,“王**”属于人物类别,确定的推荐信息例如包括推荐语句“适用于Android操作系统的开发”和推荐语句“该领域的专家有王**”。

[0114] 根据本公开的实施例,为了进一步地丰富推荐信息,在根据至少一个节点所针对的实体词确定推荐信息时,可以先根据至少一个节点对实体词进行扩充,然后根据扩充得到的实体词及至少一个节点所针对的实体词来生成推荐信息。

[0115] 示例性地,如图4B所示,在至少一个节点包括节点403、节点404、节点405和节点406时,可以先根据与至少一个节点相关联的边,确定与至少一个节点连接的节点,以对实体词进行一次扩充。例如,与至少一个节点连接的节点(图4B中右下角的阴影覆盖的节点)包括节点407、节点408和节点409。然后根据与至少一个节点连接的节点所针对的实体词和至少一个节点所针对的实体词来生成推荐信息。

[0116] 示例性地,如图4B所示,本公开实施例还可以根据至少一个节点对实体词进行二次扩充或更多次扩充,然后再根据扩充后的实体词生成推荐信息。例如,扩充得到的实体词包括:节点407~节点409和节点410~节点411所针对的实体词。最后,根据扩充得到的实体词和至少一个节点所针对的实体词生成推荐信息。

[0117] 根据本公开的实施例,为了使得知识图谱能够指示更丰富的信息,本公开实施例

还可以根据增量数据对知识图谱中的节点和边进行补充。

[0118] 图5示意性示出了根据本公开实施例的更新知识图谱的流程图。

[0119] 如图5所示,该实施例的信息推荐方法500除了前文描述的操作S210~操作S240,及操作S350外,还可以包括操作S560~操作S580,以此实现对知识图谱的更新。

[0120] 在操作S560,响应于获取到新文本,从新文本中提取表示实体的多个新实体词及多个新实体词彼此之间的关联关系。

[0121] 根据本公开的实施例,新文本例如可以是在服务器预构建知识图谱后,终端设备实时上传的文本。该操作S560与前述图3A描述的操作S351的实现方法相同或相似,在此不再赘述。

[0122] 在操作S570,判断知识图谱是否能够指示多个新实体词中的每个新实体词。

[0123] 根据本公开的实施例,可以将从新文本中提取的多个新实体词分别与知识图谱中每个节点所针对的实体词进行比对,确定知识图谱中多个节点所针对的多个实体词中是否包括多个新实体词中的每个新实体词。在确定多个节点针对的多个实体词中不包括多个新实体词中的某个新实体词时,确定知识图谱不能指示该某个新实体词。

[0124] 在知识图谱不能指示多个新实体词中的每个新实体词的情况下,即在知识图谱不能指示多个新实体词中至少一个新实体词的情况下,执行操作S580,根据多个新实体词及多个新实体词彼此之间的关联关系,更新知识图谱。在知识图谱能指示多个新实体词中的每个新实体词时,则结束操作,不对知识图谱进行更新。

[0125] 示例性地,可以先构建针对至少一个新实体词的节点,然后根据多个实体词中与至少一个新实体词中每个新实体词具有关联关系的新关联词。最后连接针对每个新实体词的节点与针对新关联词的节点,形成与针对每个新实体词的节点相关联的边。至此,完成对知识图谱的更新。

[0126] 示例性地,在知识图谱能够指示多个新实体词中的每个实体词的情况下,本公开实施例还可以根据知识图谱中的边,来确定知识图谱是否能够指示多个新实体词彼此之间的关联关系中的每个关联关系。在知识图谱不能指示多个新实体词彼此之间的关联关系中的部分关联关系时,则根据该部分关联关系在知识图谱中增加与针对新实体词的节点相关联的边。例如,若知识图谱不能指示新实体词“Python”与“脚本”之间的关联关系,但知识图谱能指示该两个新实体词,则可以通过在知识图谱中增加连接针对新实体词“Python”的节点与针对新实体词“脚本”的节点之间的边,实现对知识图谱的更新。

[0127] 为了进一步丰富向用户推荐的信息,并实现对企业内部知识的智能化管理,本公开实施例可以对用于构建知识图谱的文本进行分类化管理。示例性地,本公开实施例可以预先构建有包括多个信息节点的关键信息结构。在接收到终端设备上传的文本后,根据文本包括的关键信息来将文本归类至关键信息结构中对应的信息节点。从而,在用户请求信息时,即可根据请求信息中的关键信息来将关键信息结构的信息节点归类得到的文本作为推荐信息反馈给终端设备。

[0128] 图6A示意性示出了根据本公开另一实施例的信息推荐方法的流程图。图6B示意性示出了根据本公开实施例的预设关键信息结构的示意图。

[0129] 如图6A所示,该实施例的信息推荐方法600除了操作S210~操作S240,及操作S350外,例如还可以包括操作S660~操作S670。示例性地,该操作S660~670例如可以与操作

S220同步执行,或者可以在操作S220之前或之后执行,本公开对此不作限定。

[0130] 在操作S660,从请求信息中提取关键信息。

[0131] 根据本公开的实施例,例如可以通过对query语句进行关键词提取来得到关键信息。提取关键词可以采用TF-IDF(Term Frequency-Inverse Document Frequency,词频-逆向文本频率)算法、TextRank算法、LSA(Latent Semantic Analysis,潜在语义分析)算法、LSI(Latent Semantic Index,潜在语义索引)算法等。

[0132] 示例性地,本公开实施例可以采用TextRank算法来提取关键信息。首先,例如可以采用与操作S220类似的方式,先对请求信息进行分词处理,得到指示请求信息的词序列。然后对词序列中的每个词进行词性标注处理,并过滤掉停用词,只保留指定词性的词,例如名词、动词和形容词等。将保留的词作为候选关键信息。然后构建候选关键信息图 $G=(V,E)$,其中V为节点集,由候选关键信息组成,然后采用两个候选关键信息之间的共现关系(Co-Occurrence)来构造节点集中任两个节点之间的边。然后根据TextRank的公式,迭代传播节点集中各节点的权重,直至收敛。随后对节点集中的节点根据权重倒序排序,得到最重要的若干个词,作为从文本中提取的关键信息。

[0133] 在操作S670,确定预设关键信息结构中与所述提取的关键信息相关联的信息节点,作为目标信息节点,关键信息结构包括多个信息节点,多个信息节点中的每个信息节点用于指示一个关键信息。

[0134] 根据本公开的实施例,如图6B所示,关键信息结构例如可以为多层结构,每层结构包括至少一个信息节点,位于不同层的信息节点所指示的关键信息是属于不同粒度的信息。例如,每层结构的信息节点指示的关键信息的粒度由粗到细排列。示例性地,关键信息结构可以为三层结构,顶层结构包括信息节点601,用于指示关键信息“人工智能”。第二层结构包括信息节点602和信息节点603,该两个信息节点分别用于指示关键信息“神经网络”和“深度学习”。第三层结构包括信息节点604、信息节点605、信息节点606和信息节点607,该四个信息节点分别用于指示关键信息“CNN(卷积神经网络)”、“RNN(循环神经网络)”、“BM(Boltzmann Machines,玻尔兹曼机)”和“RBF(Radial Basis Function,径向基函数网络)”。其中,信息节点604和信息节点605属于信息节点602的细粒度的分支节点,信息节点606和信息节点607属于信息节点603的细粒度的分支节点。该关键信息结构中的每个信息节点例如可以对应有一个存储空间,用于存储被归类至该每个信息节点的文本。

[0135] 根据本公开的实施例,例如可以先确定所提取的关键信息与关键信息结构中的每个信息节点指示的关键信息的相似度,确定指示相似度高的关键信息的信息节点为目标信息节点。

[0136] 示例性地,该操作S670例如可以通过后续图7或图8描述的流程来实现,在此不再赘述。

[0137] 基于该目标信息节点的确定,前述确定推荐信息的操作S240例如可以通过图6A描述的操作S640来实现。在操作S640,根据请求实体词、预构建的知识图谱以及目标知识节点,确定推荐信息。

[0138] 根据本公开的实施例,该操作S640例如不仅可以根据权利要求实体词与预构建的知识图谱通过图6A描述的流程确定推荐信息,还可以根据目标知识节点来确定推荐文本,并将推荐文本作为推荐信息。

[0139] 示例性地,根据目标知识节点来确定推荐文本可以包括:获取归类至目标知识节点的文本,并将该获取的文本作为推荐信息。

[0140] 根据本公开的实施例,为了便于向用户推荐文本,本公开实施例的信息推荐方法例如还可以在基于文本构建知识图谱时,将文本归类为关键信息结构包括的信息节点。以此可以便于根据信息节点来向用户推荐文本,并同时实现文本的分类管理。前述通过操作S351提取的多个实体词中例如可以包括有用户表示文本的关键信息的第二词,该第二词例如可以为通过类似于操作S660的方法从文本中提取得到的关键词。

[0141] 图7示意性示出了根据本公开实施例的将文本归类为关键信息结构包括的信息节点的流程图。

[0142] 如图7所示,将文本归类为关键信息结构包括的信息节点的操作S780例如可以包括操作S781~操作S782。

[0143] 在操作S781,根据第二词,确定文本与关键信息结构中每个信息节点之间的匹配关系。在操作S782,将文本归类为与文本相匹配的信息节点。

[0144] 根据本公开的实施例,可以将多个实体词中的第二词与每个信息节点指示的关键信息之间的相似关系,来确定文本与每个信息节点之间的匹配关系。根据匹配关系来确定与文本相匹配的信息节点,并将文本存储至相匹配的信息节点所对应的存储空间。

[0145] 示例性地,操作S781例如可以通过后续图8描述的流程或图9描述的流程来实现,在此不再详述。

[0146] 根据本公开的实施例,类似于前述的知识图谱,预设的关键信息结构例如也可以根据增量数据进行补充。例如,该实施例的信息推荐方法可以响应于获取到新文本,先执行从新文本中提取新文本的关键信息的操作,并将提取的关键词作为新关键信息。该操作的实现方式与前述操作S660的实现方式类似,在此不再赘述。在提取到新关键信息后,再判断关键信息结构中是否包括指示该新关键信息的信息节点。在不包括的情况下,根据新关键信息,更新关键信息结构。

[0147] 示例性地,可以根据新关键信息与关键信息结构中多个信息节点指示的关键信息的关联关系,增加针对新关键信息的信息节点。例如,若新关键信息为“LSTM”,则根据新关键信息“LSTM”与图6B中信息节点605指示的关键信息“RNN”的从属关系,在信息节点605的下层结构中添加属于信息节点605的分支节点,作为针对“LSTM”的信息节点。

[0148] 图8示意性示出了根据本公开实施例的确定文本与关键信息结构中每个信息节点之间的匹配关系的流程图。

[0149] 如图8所示,在一实施例中,确定文本与关键信息结构中每个信息节点之间的匹配关系的操作S781可以包括从操作S8811开始的操作S8811~操作S8812。

[0150] 在操作S8811,判断第二词是否包括目标词,目标词与每个信息节点指示的关键信息相同。

[0151] 根据本公开的实施例,可以将第二词与关键信息结构中的每个信息节点指示的关键信息进行比对,若第二词与关键信息一致,则确定第二词为目标词。

[0152] 示例性地,从文本中提取的实体词例如可以包括多个第二词,该实施例可以是多个第二词中的每个第二词与关键信息结构中的每个信息节点指示的关键信息进行比对,来确定第二词是否包括目标词。

[0153] 在第二词包括目标词的情况下,执行操作S8812,确定文本与每个信息节点之间的匹配关系为相匹配。

[0154] 考虑到对于不同场景,相同的词的表述方式可能不同,则为了避免因关键信息结构中包括的信息节点指示的词的表述方式与第二词的表述方式不同导致的无法准确确定文本与信息节点的匹配关系。该实施例的确定文本与关键信息结构中每个信息节点之间的匹配关系的操作除了操作S8811~操作S8812外,还可以包括操作S8813~操作S8815。相应地,知识图谱中例如可以具有针对同一词的不同表述方式的节点。示例性地,知识图谱中的节点不仅包括针对实体“CNN”的节点,还包括针对实体“卷积神经网络”的节点。

[0155] 在第二词不包括目标词的情况下,执行操作S8813,从知识图谱中获取第二词的关联词。示例性地,可以先确定知识图谱中用于指示第二词的节点。然后根据与指示第二词的节点相关联的边来确定第二词的关联词。例如,将相关联的边所连接的节点指示的实体词作为关联词。

[0156] 在操作S8814,判断第二词的关联词是否包括目标词。该操作S8814的实现方法与操作S8811的实现方法类似,在此不再赘述。

[0157] 在第二词的关联词包括目标词的情况下,执行操作S8812,确定文本与每个信息节点之间的匹配关系为相匹配。

[0158] 在第二词的关联词不包括目标词的情况下,则执行操作S8815,确定文本与每个信息节点之间的匹配关系为不匹配。

[0159] 根据本公开的实施例,在确定文本与信息节点的匹配关系时,通过基于知识图谱来获取文本中的第二词的关联词,可以使得进行匹配时采用的词能够更全面的提现文本的内容,从而提高文本与信息节点匹配关系确定的准确性。

[0160] 图9示意性示出了根据本公开另一实施例的确定文本与关键信息结构中每个信息节点之间的匹配关系的流程图。

[0161] 如图9所示,在另一实施例中,确定文本与关键信息结构中每个信息节点之间的匹配关系的操作S781可以包括操作S9811~操作S9814。

[0162] 在操作S9811,确定第二词与每个信息节点指示的关键信息之间的相似度。

[0163] 根据本公开的实施例,该操作例如可以通过计算第二词与每个信息节点指示的关键信息之间的余弦相似度、杰卡德相似系数来得到两者之间的相似度。示例性地,在第二词的数量为多个时,该操作S9811为确定每个第二词与每个信息节点指示的关键信息之间的相似度,得到多个相似度。

[0164] 在操作S9812,判断相似度是否大于相似度阈值。其中,相似度阈值例如可以根据实际需求进行设定,本公开对此不作限定。是理性地,该相似度阈值可以取0.9。

[0165] 在相似度大于相似度阈值的情况下,执行操作S9813,确定文本与每个信息节点之间的匹配关系为相匹配。在相似度不大于相似度阈值的情况下,执行操作S9814,确定文本与每个信息节点之间的匹配关系为不匹配。

[0166] 示例性地,在第二词的数量为多个时,该操作S9812可以为判断多个相似度中是否包括大于相似度阈值的相似度。若包括,则执行操作S9813,若不包括,则执行操作S9814。

[0167] 通过前述流程,本公开实施例可以先从文本中提取关键词,然后根据关键词确定关键信息结构中匹配的信息节点。最后将文本存储至匹配的信息节点所对应的存储空间。

在从文本中提取得到多个关键词时,可以从关键信息结构中得到多个匹配的信息节点。此种情况,则将文本复制为多份,存储至多个匹配的信息节点对应的存储空间。

[0168] 示例性地,为了避免占用过多存储空间,可以为每个信息节点维护一个文本列表。该文本列表包括可以归类为该信息节点的文本的题录等。在确定推荐信息后向终端设备推送文本时,可以根据信息节点的文本列表获取文本。

[0169] 可以理解的是,本公开实施例提供的信息推荐方法可以应用于多种应用场景。例如,该信息推荐方法可以应用于以下至少一个应用场景:进行知识搜索的应用场景,进行知识问答的应用场景,和进行信息主动推荐的应用场景。

[0170] 图10示意性示出了根据本公开实施例的用于实现信息推荐方法的系统架构图。

[0171] 如图10所示,该实施例的用于实现信息推荐方法的系统架构1000例如可以包括知识预处理模块1010、知识图谱构建模块1020、知识服务模块1030和知识应用模块1040。

[0172] 知识预处理模块例如可以用于先通过前述图3B描述的流程来对作为知识的信息进行收集,使得各种信息形成在线知识。然后对形成的在线知识进行知识清洗,仅保留知识性强、内容质量高的线上知识。

[0173] 知识图谱构建模块1020例如可以用于通过执行图3A描述的流程来构建知识图谱。例如,先对知识预处理模块1010得到的知识进行挖掘。通过知识挖掘,例如可以得到在线知识的各种变形。示例性地,若在线知识为中文表述的知识,通过知识挖掘,例如可以得到英文表述的知识。在对知识进行挖掘后,对在线知识和挖掘得到的知识进行实体挖掘,即从知识中提取多个实体词及多个实体词彼此之间的关联关系。最后对实体挖掘得到的信息进行图谱化,得到知识图谱。图谱化过程例如可以通过前述操作S352来实现。

[0174] 示例性地,知识图谱构建模块1020例如还可以用于构建知识分类体系,以实现知识的归类。例如,知识图谱构建模块1020可以用于通过图7描述的流程来对作为知识的文本进行分类,实现知识分类体系的构建。

[0175] 知识服务模块1030例如可以先基于知识图谱,生成向用户推荐的知识中的知识元素。生成的知识元素可以包括通过前述图4A描述的流程得到的实体词。随后对生成的知识元素进行聚合,例如可以通过前述操作S433进行知识聚合。同时,知识服务模块1030还可以从知识分类体系中检索知识,知识的检索例如可以是图6A描述的流程来完成,并通过检索得到归类为目标信息节点的文本。需要说明的是,在知识检索的过程中,若不能从知识分类体系中得到文本,则可以在进行知识推理后再次进行知识检索。知识推理例如可以通过图8描述的操作S8813来完成。

[0176] 基于知识服务模块1030的功能,知识应用模块1040可以通过终端设备向用户提供“知识门户”功能模块、“知识检索”功能模块、“知识问答”功能模块和“场景化推荐”功能模块。其中,通过“知识门户”功能模块,用户例如可以进行知识的上传和下载等。通过“知识检索”功能模块,用户例如可以输入检索语句,来从知识图谱和知识分类体系中检索到想要了解的知识。通过“知识问答”功能模块,用户例如可以输入询问语句,来从知识图谱和知识分类体系中获取到能够解惑的知识。通过“场景化推荐”功能模块,终端设备例如可以根据用户喜好向用户推荐适合用户的知识。

[0177] 综上所述,本公开实施例通过适于执行信息推荐方法的系统架构的构建,可以为企业的知识智能化管理与应用提供完整的解决方案。将企业内分散、碎片化的文档进行有

效的筛选与聚合,实现知识的体系化管理和企业知识图谱的构建,从而进一步将企业知识图谱与工作场景充分融合,实现工作场景知识的智能化应用,提升企业知识生产和应用的效率。

[0178] 图11示意性示出了根据本公开实施例的信息推荐装置的结构框图。

[0179] 如图11所示,该实施例的信息推荐装置1100例如可以包括获取模块1110、信息提取模块1120、信息确定模块1130和信息推送模块1140。

[0180] 获取模块1110用于获取请求信息。在一实施例中,获取模块1110可以用于执行图2描述的操作S210,在此不再赘述。

[0181] 信息提取模块1120用于从请求信息中提取表示实体的请求实体词。在一实施例中,信息提取模块1120可以用于执行图2描述的操作S220,在此不再赘述。

[0182] 信息确定模块1130用于根据请求实体词和预构建的知识图谱,确定推荐信息。其中,知识图谱是基于文本构建的,且知识图谱指示了对文本的来源加以表示的第一词。在一实施例中,信息确定模块1130可以用于执行图2描述的操作S230,在此不再赘述。

[0183] 信息推送模块1140用于推送推荐信息。在一实施例中,信息推送模块1140可以用于执行图2描述的操作S240,在此不再赘述。

[0184] 根据本公开的实施例,信息推荐装置1100例如还可以包括图谱构建模块,用于基于文本构建知识图谱。例如,该图谱构建模块例如可以用于执行图3A和图3C描述的流程,在此不再赘述。

[0185] 根据本公开的实施例,前述信息确定模块1130例如可以用于执行图4A描述的流程和/或图6A描述的操作S640,在此不再赘述。

[0186] 根据本公开的实施例,前述信息提取模块1120例如还可以用于执行图5描述的操作S560。信息推荐装置1100例如还可以包括判断模块和更新模块。判断模块例如可以用于执行图5描述的操作S570,更新模块例如可以用于执行图5描述的操作S580,在此不再赘述。

[0187] 根据本公开的实施例,前述信息提取模块1120例如还可以用于执行图6A描述的操作S660。信息推荐装置1100例如还可以包括节点确定模块,用于执行图6A描述的操作S670,在此不再赘述。

[0188] 根据本公开的实施例,信息推荐装置1100例如还可以包括文本归类模块,用于将文本归类为关键信息结构包括的信息节点。该文本归类模块例如可以用于执行图7描述的操作S781~操作S782,在此不再赘述。

[0189] 根据本公开的实施例,文本归类模块例如可以通过前述图8描述的流程或图9描述的流程来确定文本与关键信息结构中每个信息节点之间的匹配关系,在此不再赘述。

[0190] 根据本公开的实施例的模块、子模块、单元、子单元中的任意多个、或其中任意多个的至少部分功能可以在一个模块中实现。根据本公开实施例的模块、子模块、单元、子单元中的任意一个或多个可以被拆分成多个模块来实现。根据本公开实施例的模块、子模块、单元、子单元中的任意一个或多个可以至少被部分地实现为硬件电路,例如现场可编程门阵列(FPGA)、可编程逻辑阵列(PLA)、片上系统、基板上的系统、封装上的系统、专用集成电路(ASIC),或可以通过对电路进行集成或封装的任何其他的合理方式的硬件或固件来实现,或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者,根据本公开实施例的模块、子模块、单元、子单元中的一个或多个可以至少被部分

地实现为计算机程序模块,当该计算机程序模块被运行时,可以执行相应的功能。

[0191] 根据本公开的实施例,本公开还提供了一种计算机系统、一种可读存储介质和一种计算机程序产品。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可以实现上述任意实施例的方法。

[0192] 图12示意性示出了根据本公开实施例的适于执行信息推荐方法的计算机系统的方框图。

[0193] 如图12所示,根据本公开实施例的计算机系统1200包括处理器1201,其可以根据存储在只读存储器 (ROM) 1202中的程序或者从存储部分1208加载到随机访问存储器 (RAM) 1203中的程序而执行各种适当的动作和处理。处理器1201例如可以包括通用微处理器(例如CPU)、指令集处理器和/或相关芯片组和/或专用微处理器(例如,专用集成电路(ASIC)),等等。处理器1201还可以包括用于缓存用途的板载存储器。处理器1201可以包括用于执行根据本公开实施例的方法流程的不同动作的单一处理单元或者是多个处理单元。

[0194] 在RAM 1203中,存储有系统1200操作所需的各种程序和数据。处理器1201、ROM 1202以及RAM 1203通过总线1204彼此相连。处理器1201通过执行ROM 1202和/或RAM 1203中的程序来执行根据本公开实施例的方法流程的各种操作。需要注意,所述程序也可以存储在除ROM 1202和RAM 1203以外的一个或多个存储器中。处理器1201也可以通过执行存储在所述一个或多个存储器中的程序来执行根据本公开实施例的方法流程的各种操作。

[0195] 根据本公开的实施例,系统1200还可以包括输入/输出 (I/O) 接口1205,输入/输出 (I/O) 接口1205也连接至总线1204。系统1200还可以包括连接至I/O接口1205的以下部件中的一项或多项:包括键盘、鼠标等的输入部分1206;包括诸如阴极射线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分1207;包括硬盘等的存储部分1208;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分1209。通信部分1209经由诸如因特网的网络执行通信处理。驱动器1210也根据需要连接至I/O接口1205。可拆卸介质1211,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器1210上,以便于从其上读出的计算机程序根据需要被安装入存储部分1208。

[0196] 根据本公开的实施例,根据本公开实施例的方法流程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读存储介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分1209从网络上被下载和安装,和/或从可拆卸介质1211被安装。在该计算机程序被处理器1201执行时,执行本公开实施例的系统中限定的上述功能。根据本公开的实施例,上文描述的系统、设备、装置、模块、单元等可以通过计算机程序模块来实现。

[0197] 本公开还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备/装置/系统中所包含的;也可以是单独存在,而未装配入该设备/装置/系统中。上述计算机可读存储介质承载有一个或者多个程序,当上述一个或者多个程序被执行时,实现根据本公开实施例的方法。

[0198] 根据本公开的实施例,计算机可读存储介质可以是非易失性的计算机可读存储介质,例如可以包括但不限于:便携式计算机磁盘、硬盘、随机访问存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (EPROM或闪存)、便携式紧凑磁盘只读存储器 (CD-ROM)、光

存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。例如,根据本公开的实施例,计算机可读存储介质可以包括上文描述的ROM 1202和/或RAM 1203和/或ROM 1202和RAM 1203以外的一个或多个存储器。

[0199] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0200] 本领域技术人员可以理解,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合,即使这样的组合或结合没有明确记载于本公开中。特别地,在不脱离本公开精神和教导的情况下,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合。所有这些组合和/或结合均落入本公开的范围。

[0201] 以上对本公开的实施例进行了描述。但是,这些实施例仅仅是为了说明的目的,而并非为了限制本公开的范围。尽管在以上分别描述了各实施例,但是这并不意味着各个实施例中的措施不能有利地结合使用。本公开的范围由所附权利要求及其等同物限定。不脱离本公开的范围,本领域技术人员可以做出多种替代和修改,这些替代和修改都应落在本公开的范围之内。

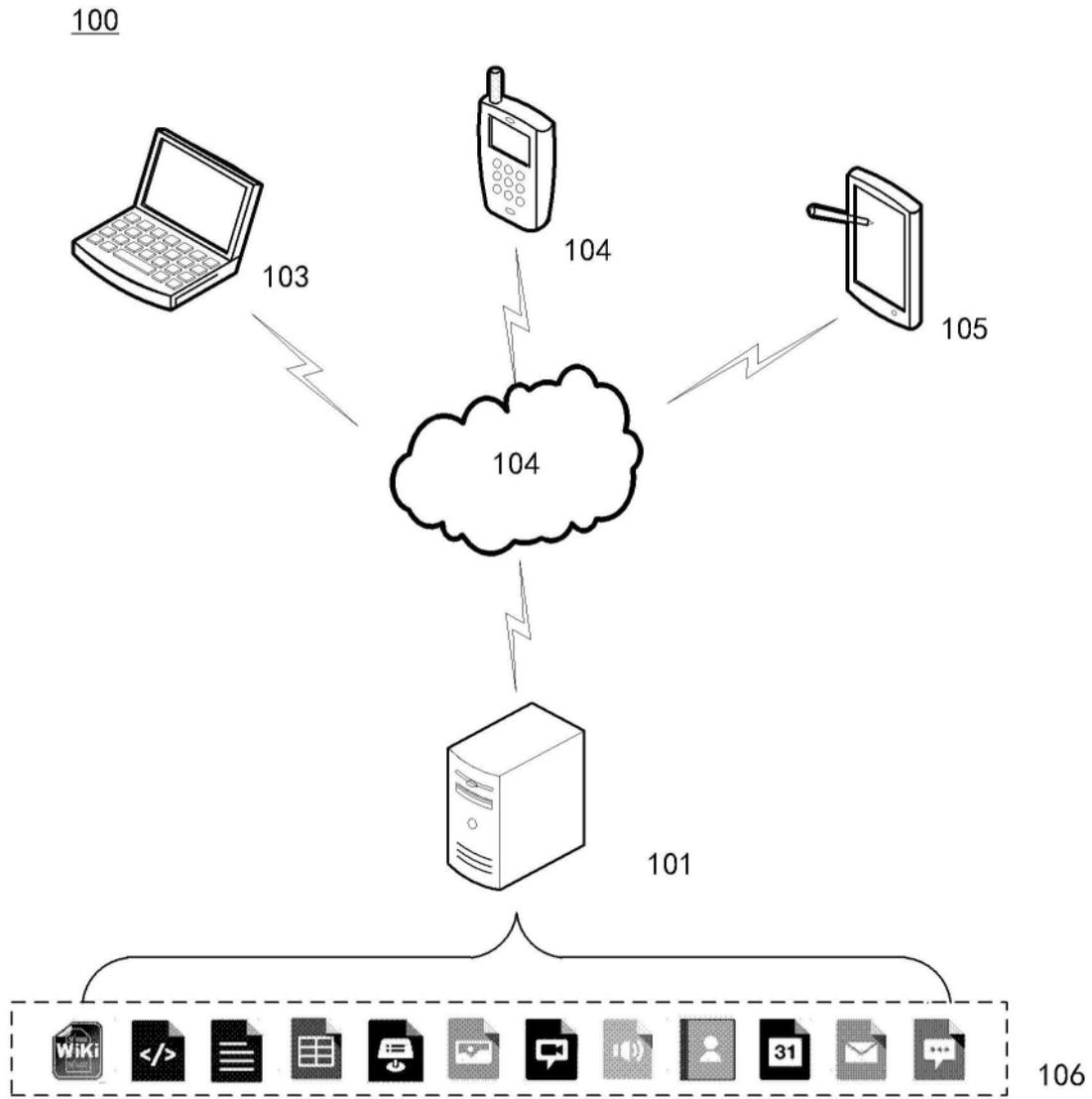


图1

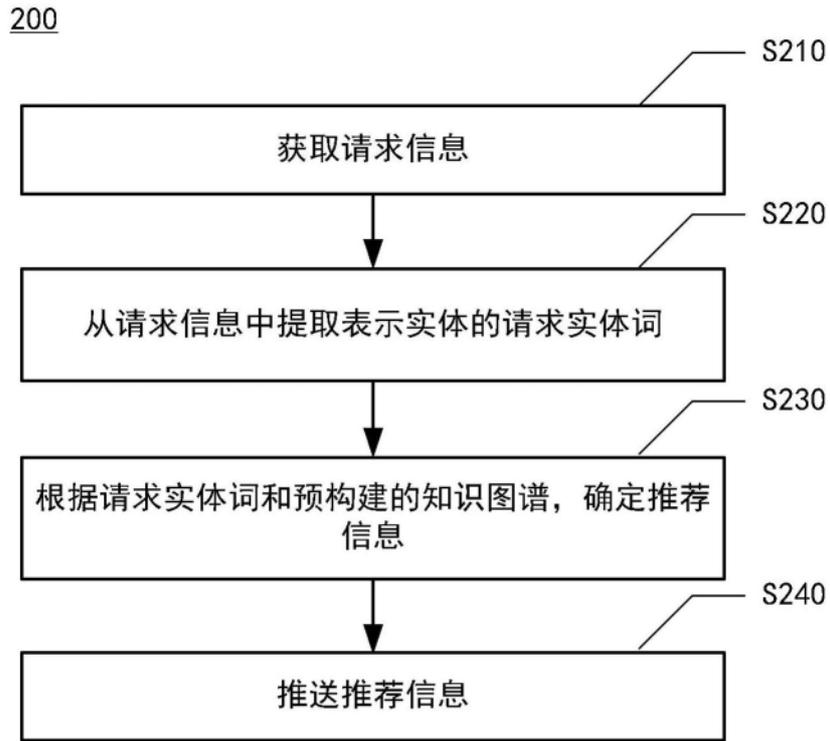


图2

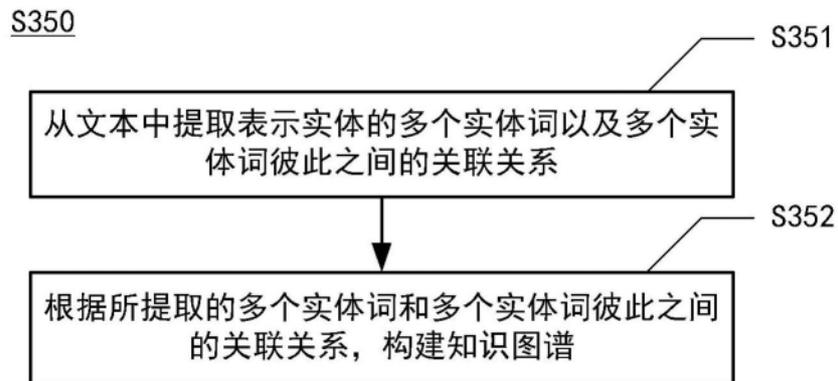


图3A

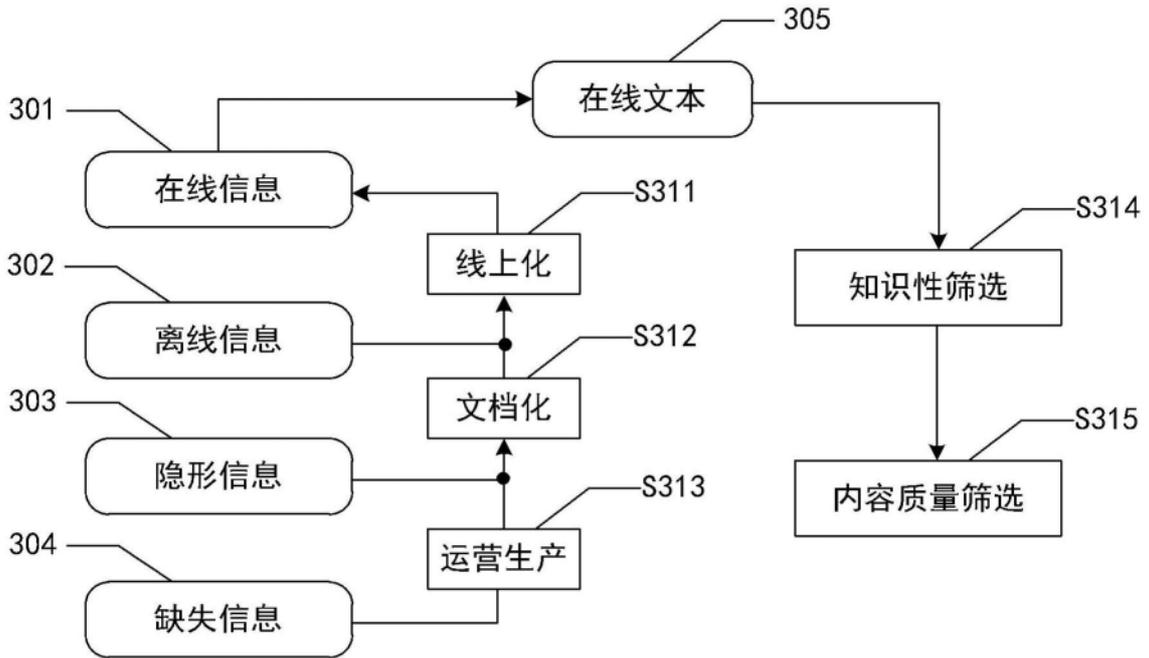


图3B

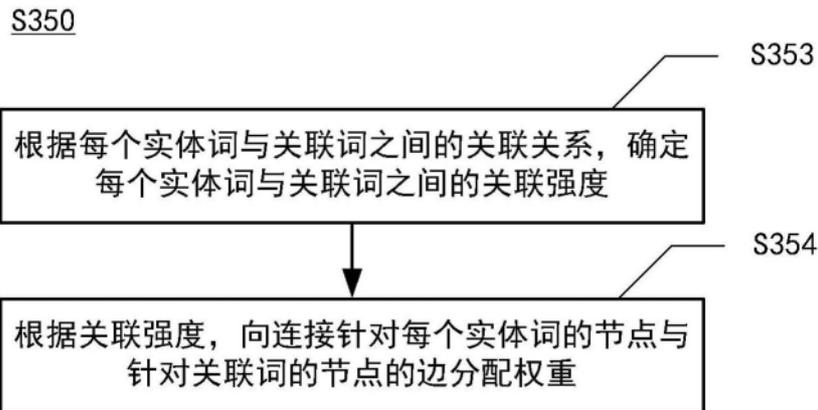


图3C

S230

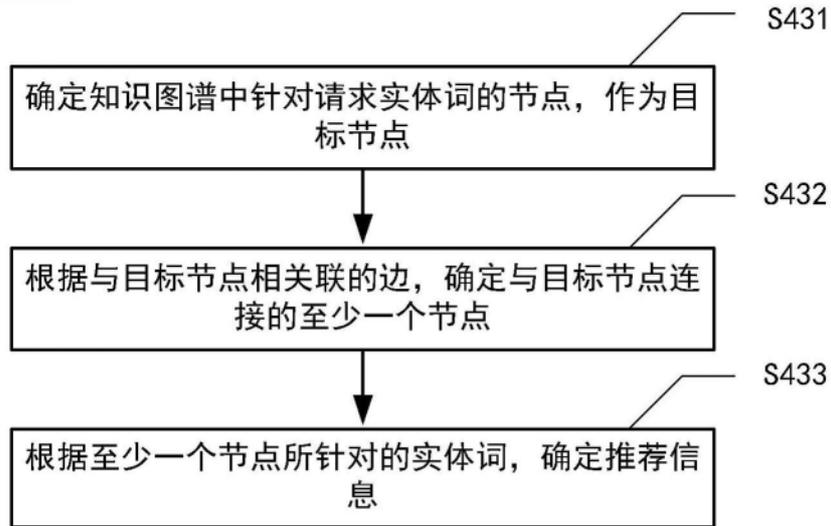


图4A

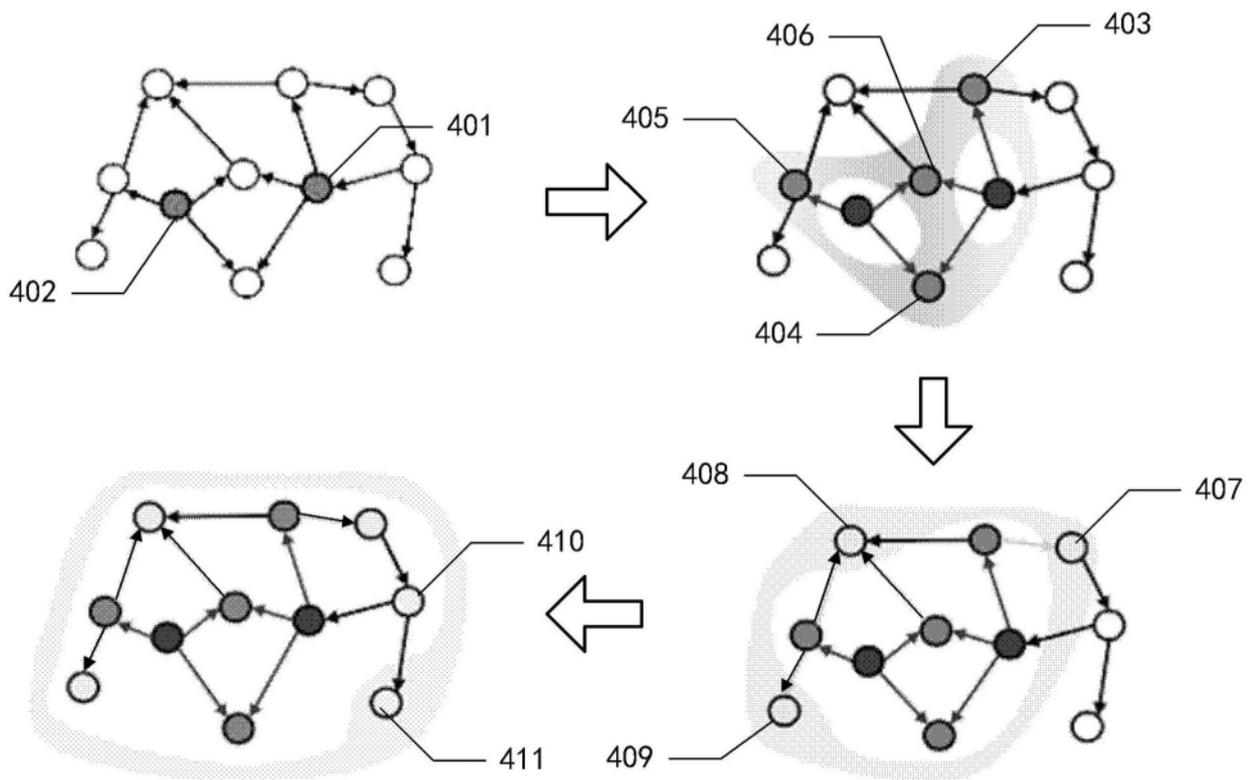


图4B

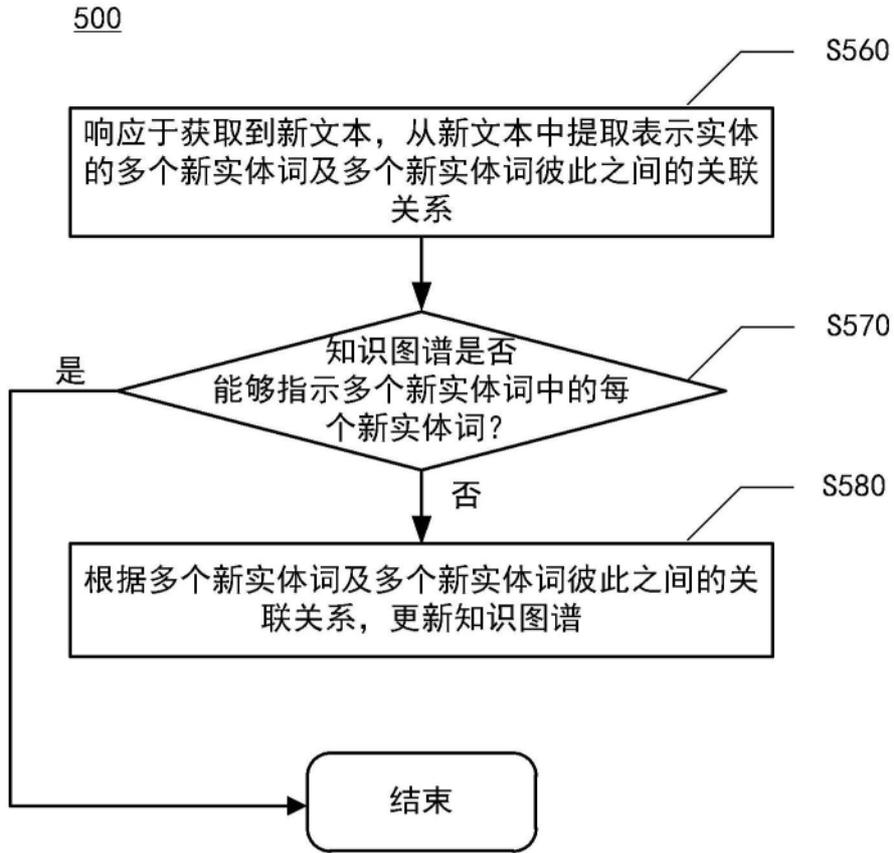


图5

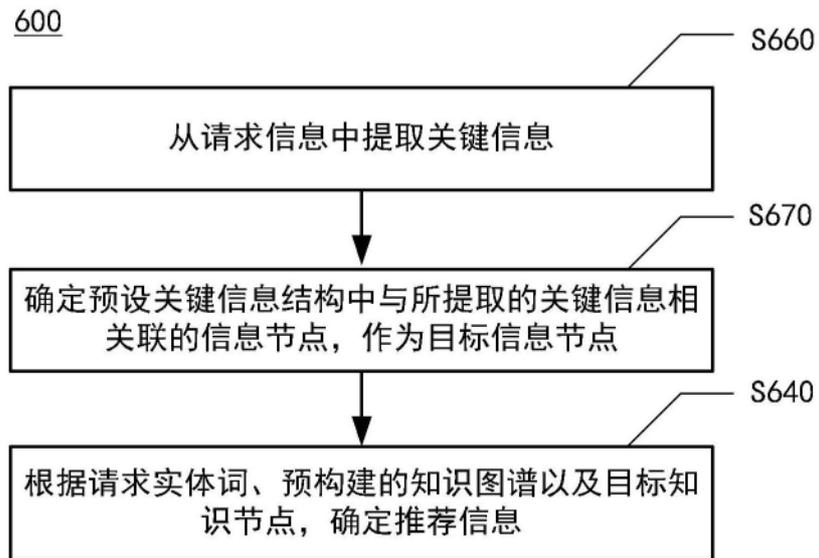


图6A

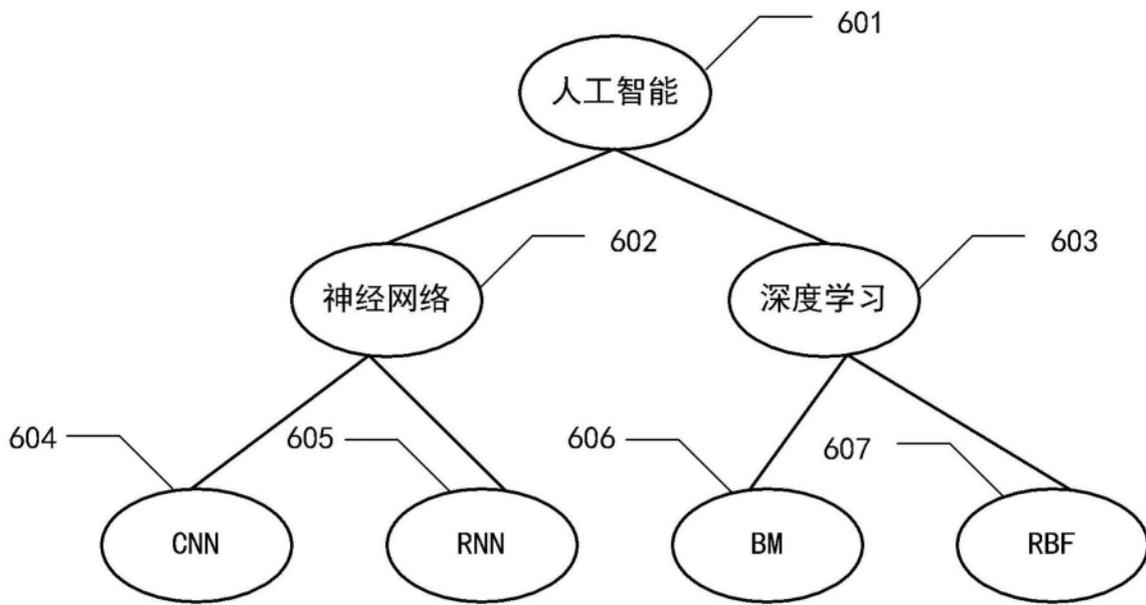


图6B

S780

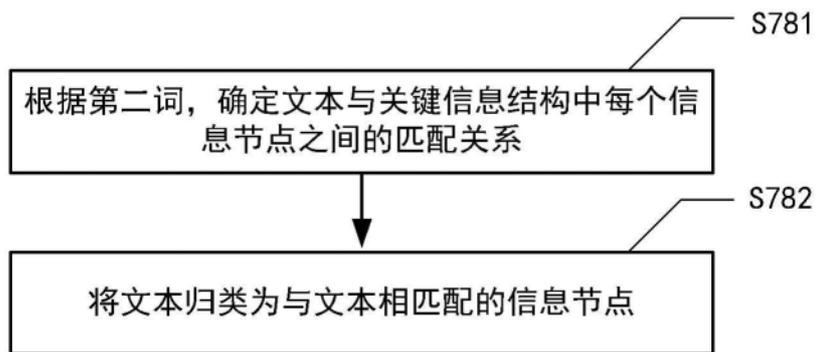


图7

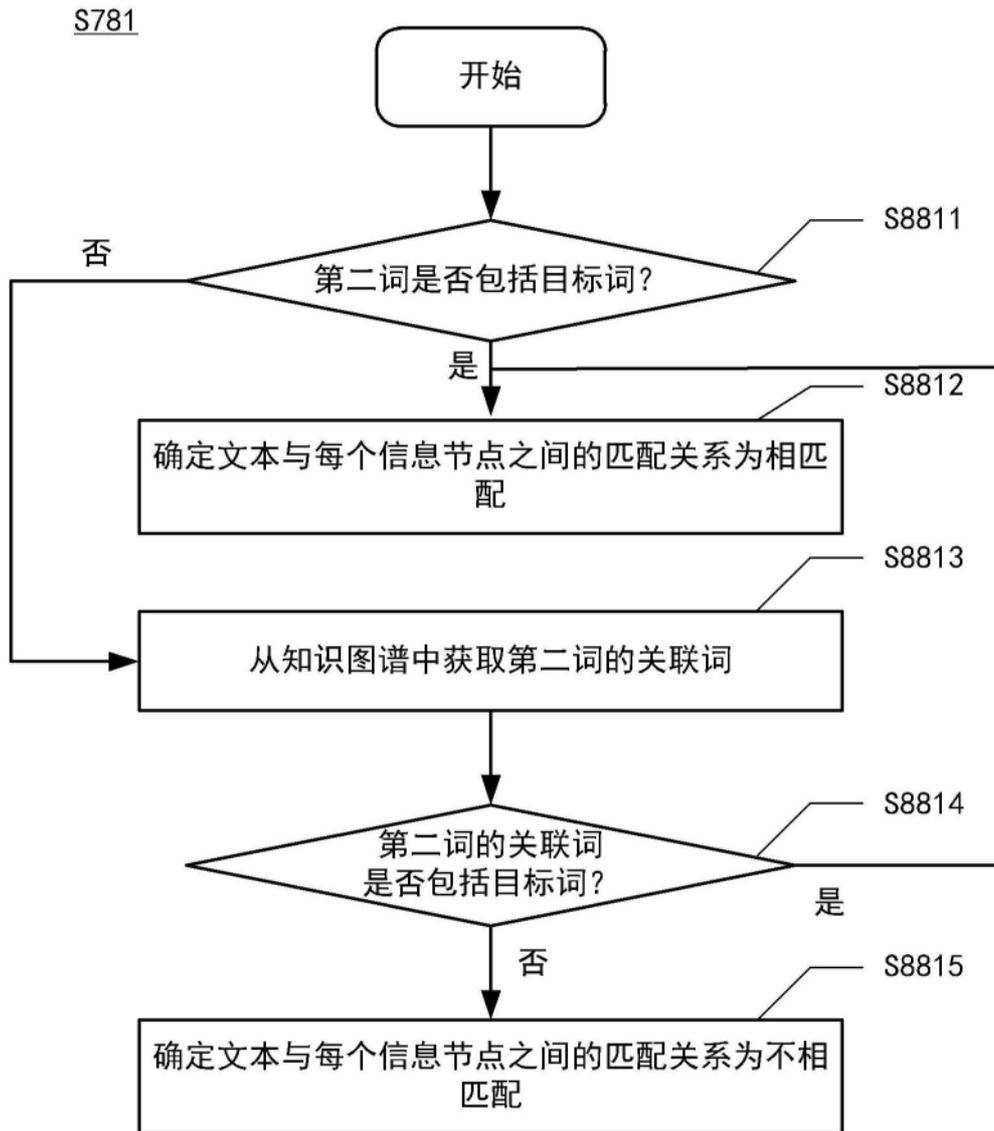


图8

S781

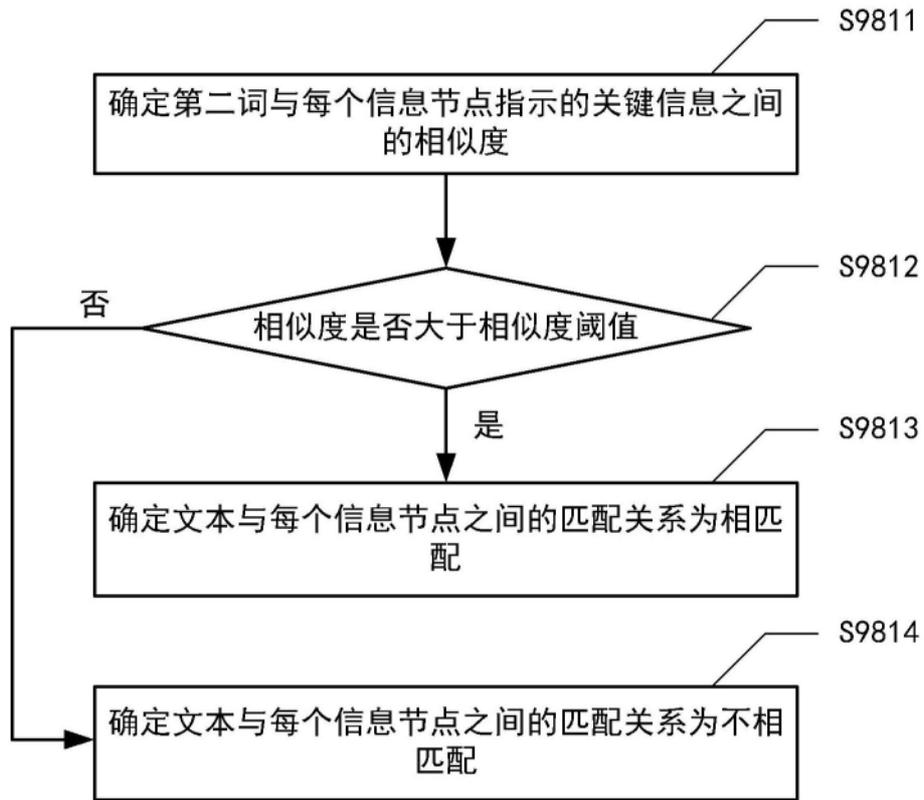


图9

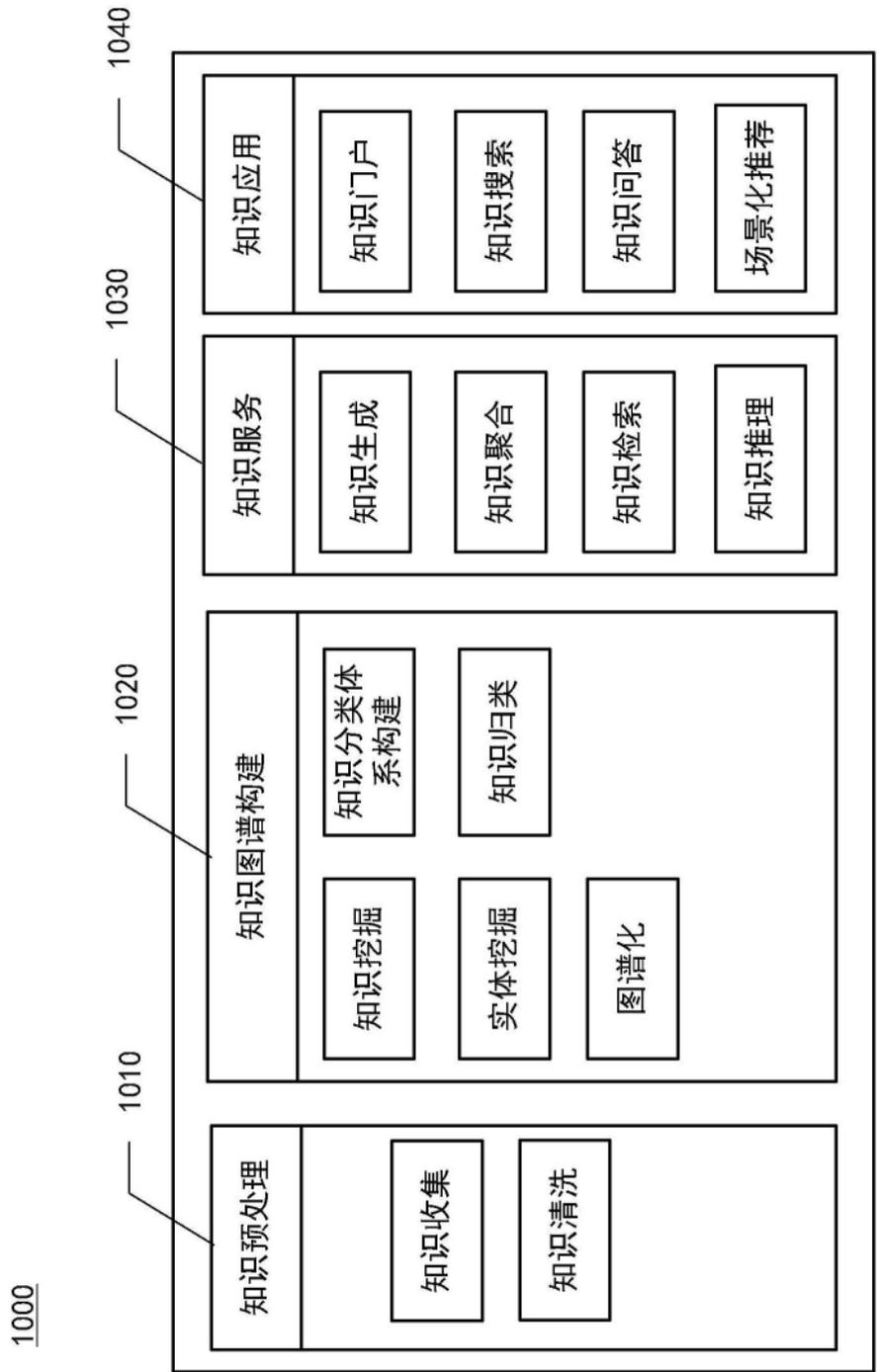


图10

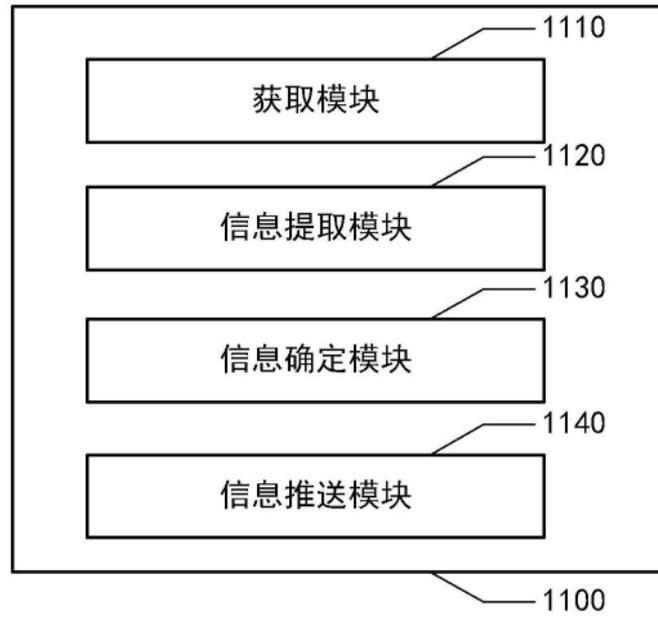


图11

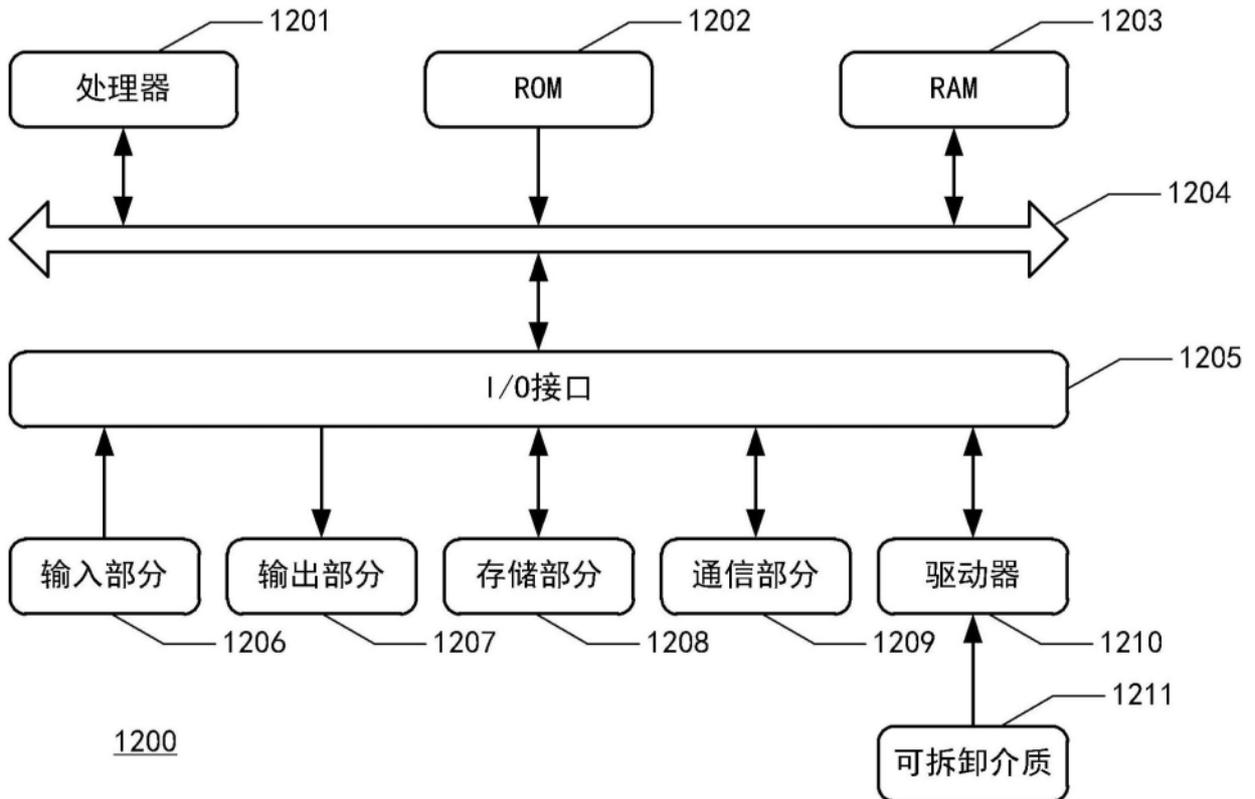


图12