



(12)发明专利申请

(10)申请公布号 CN 109165344 A

(43)申请公布日 2019.01.08

(21)申请号 201810887222.8

(22)申请日 2018.08.06

(71)申请人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街10号

百度大厦三层

(72)发明人 张丰

(74)专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

代理人 王达佐 马晓亚

(51) Int. Cl.

G06F 16/9535(2019.01)

G06F 16/332(2019.01)

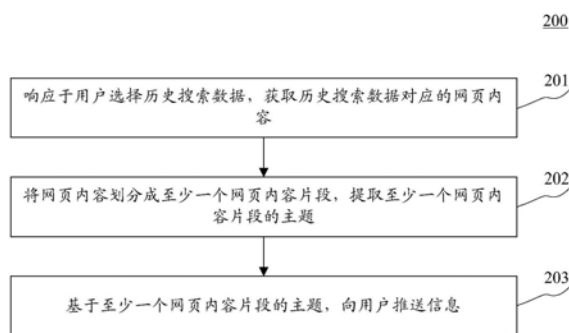
权利要求书2页 说明书11页 附图4页

(54)发明名称

用于推送信息的方法和装置

(57)摘要

本申请实施例公开了用于推送信息的方法和装置。该方法的一具体实施方式包括：响应于用户选择历史搜索数据，获取历史搜索数据对应的网页内容；将网页内容划分成至少一个网页内容片段，提取至少一个网页内容片段的主题；基于至少一个网页内容片段的主题，向用户推送信息。该实施方式能够在用户再次浏览网页时，快速地向用户推送网页的核心信息。



1. 一种用于推送信息的方法,包括:
 - 响应于用户选择历史搜索数据,获取所述历史搜索数据对应的网页内容;
 - 将所述网页内容划分成至少一个网页内容片段,提取所述至少一个网页内容片段的主题;
 - 基于所述至少一个网页内容片段的主题,向所述用户推送信息。
2. 根据权利要求1所述的方法,其中,所述提取所述至少一个网页内容片段的主题,包括:
 - 对于所述至少一个网页内容片段中的网页内容片段,将该网页内容片段输入至预先训练的文档主题生成模型,得到该网页内容片段的主题。
3. 根据权利要求1所述的方法,其中,所述基于所述至少一个网页内容片段的主题,向所述用户推送信息,包括:
 - 将所述至少一个网页内容片段的主题推送给所述用户。
4. 根据权利要求1或3所述的方法,其中,所述基于所述至少一个网页内容片段的主题,向所述用户推送信息,包括:
 - 对于所述至少一个网页内容片段中的网页内容片段,将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字,对所述匹配的文字设置预设文字效果,生成设置文字效果的网页;
 - 将所述设置文字效果的网页推送给用户。
5. 根据权利要求1所述的方法,其中,所述基于所述至少一个网页内容片段的主题,向所述用户推送信息,包括:
 - 对所述至少一个网页内容片段的主题进行切词,基于切词结果生成所述网页内容对应的词向量;
 - 将所述网页内容对应的词向量输入至预先训练的意图预测模型,得到所述网页内容对应的搜索意图,其中,所述意图预测模型用于预测搜索意图;
 - 从所述用户的用户画像中提取兴趣标签集合;
 - 将所述网页内容对应的搜索意图在所述兴趣标签集合中进行匹配,获取匹配的兴趣标签;
 - 将所述网页内容中的文字与所述匹配的兴趣标签进行匹配,获取匹配的文字,对所述匹配的文字设置预设文字效果,生成设置文字效果的网页;
 - 将所述设置文字效果的网页推送给所述用户。
6. 根据权利要求5所述的方法,其中,所述意图预测模型通过如下步骤训练得到:
 - 获取训练样本集合,其中,训练样本包括样本搜索数据和样本搜索数据对应的样本搜索意图;
 - 基于所述训练样本集合,对多层神经网络进行训练,得到所述意图预测模型。
7. 一种用于推送信息的装置,包括:
 - 获取单元,被配置成响应于用户选择历史搜索数据,获取所述历史搜索数据对应的网页内容;
 - 提取单元,被配置成将所述网页内容划分成至少一个网页内容片段,提取所述至少一个网页内容片段的主题;

推送单元,被配置成基于所述至少一个网页内容片段的主题,向所述用户推送信息。

8. 根据权利要求7所述的装置,其中,所述提取单元进一步被配置成:

对于所述至少一个网页内容片段中的网页内容片段,将该网页内容片段输入至预先训练的文档主题生成模型,得到该网页内容片段的主题。

9. 根据权利要求7所述的装置,其中,所述推送单元包括:

第一推送子单元,被配置成将所述至少一个网页内容片段的主题推送给所述用户。

10. 根据权利要求7或9所述的装置,其中,所述推送单元包括:

第一生成子单元,被配置成对于所述至少一个网页内容片段中的网页内容片段,将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字,对所述匹配的文字设置预设文字效果,生成设置文字效果的网页;

第二推送子单元,被配置成将所述设置文字效果的网页推送给用户。

11. 根据权利要求7所述的装置,其中,所述推送单元包括:

切词子单元,被配置成对所述至少一个网页内容片段的主题进行切词,基于切词结果生成所述网页内容对应的词向量;

预测子单元,被配置成将所述网页内容对应的词向量输入至预先训练的意图预测模型,得到所述网页内容对应的搜索意图,其中,所述意图预测模型用于预测搜索意图;

提取子单元,被配置成从所述用户的用户画像中提取兴趣标签集合;

匹配子单元,被配置成将所述网页内容对应的搜索意图在所述兴趣标签集合中进行匹配,获取匹配的兴趣标签;

第二生成子单元,被配置成将所述网页内容中的文字与所述匹配的兴趣标签进行匹配,获取匹配的文字,对所述匹配的文字设置预设文字效果,生成设置文字效果的网页;

第三推送子单元,被配置成将所述设置文字效果的网页推送给所述用户。

12. 根据权利要求11所述的装置,其中,所述意图预测模型通过如下步骤训练得到:

获取训练样本集合,其中,训练样本包括样本搜索数据和样本搜索数据对应的样本搜索意图;

基于所述训练样本集合,对多层神经网络进行训练,得到所述意图预测模型。

13. 一种服务器,包括:

一个或多个处理器;

存储装置,其上存储有一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-6中任一所述的方法。

14. 一种计算机可读介质,其上存储有计算机程序,其中,所述计算机程序被处理器执行时实现如权利要求1-6中任一所述的方法。

用于推送信息的方法和装置

技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及用于推送信息的方法和装置。

背景技术

[0002] 搜索可以让用户获取海量的知识,同时带来选择困难。有时候用户浏览了许多网页后,发现已经忘记了之前浏览过的网页的内容,想要再次查看时,就需要点开之前浏览过的网页重新进行阅读。

[0003] 现有的搜索过程通常包括以下几个步骤:首先,用户输入搜索词,发起搜索请求;之后,展现与搜索词关联的网页的主题;然后,用户点击网页的主题,跳转展示网页,以供用户浏览网页的内容;最后,用户浏览完毕,关闭网页。若用户想要再次查看之前浏览过的网页的内容,通常需要点击历史浏览记录中的网页的主题,或者再次输入搜索词发起搜索,并点击展现的网页的主题,跳转展示网页,以供用户再次阅读网页的内容。

发明内容

[0004] 本申请实施例提出了用于推送信息的方法和装置。

[0005] 第一方面,本申请实施例提供了一种用于推送信息的方法,包括:响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容;将网页内容划分成至少一个网页内容片段,提取至少一个网页内容片段的主题;基于至少一个网页内容片段的主题,向用户推送信息。

[0006] 在一些实施例中,提取至少一个网页内容片段的主题,包括:对于至少一个网页内容片段中的网页内容片段,将该网页内容片段输入至预先训练的文档主题生成模型,得到该网页内容片段的主题。

[0007] 在一些实施例中,基于至少一个网页内容片段的主题,向用户推送信息,包括:将至少一个网页内容片段的主题推送给用户。

[0008] 在一些实施例中,基于至少一个网页内容片段的主题,向用户推送信息,包括:对于至少一个网页内容片段中的网页内容片段,将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字,对匹配的文字设置预设文字效果,生成设置文字效果的网页;将设置文字效果的网页推送给用户。

[0009] 在一些实施例中,基于至少一个网页内容片段的主题,向用户推送信息,包括:对至少一个网页内容片段的主题进行切词,基于切词结果生成网页内容对应的词向量;将网页内容对应的词向量输入至预先训练的意图预测模型,得到网页内容对应的搜索意图,其中,意图预测模型用于预测搜索意图;从用户的用户画像中提取兴趣标签集合;将网页内容对应的搜索意图在兴趣标签集合中进行匹配,获取匹配的兴趣标签;将网页内容中的文字与匹配的兴趣标签进行匹配,获取匹配的文字,对匹配的文字设置预设文字效果,生成设置文字效果的网页;将设置文字效果的网页推送给用户。

[0010] 在一些实施例中,意图预测模型通过如下步骤训练得到:获取训练样本集合,其

中,训练样本包括样本搜索数据和样本搜索数据对应的样本搜索意图;基于训练样本集合,对多层神经网络进行训练,得到意图预测模型。

[0011] 第二方面,本申请实施例提供了一种用于推送信息的装置,包括:获取单元,被配置成响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容;提取单元,被配置成将网页内容划分成至少一个网页内容片段,提取至少一个网页内容片段的主题;推送单元,被配置成基于至少一个网页内容片段的主题,向用户推送信息。

[0012] 在一些实施例中,提取单元进一步被配置成:对于至少一个网页内容片段中的网页内容片段,将该网页内容片段输入至预先训练的文档主题生成模型,得到该网页内容片段的主题。

[0013] 在一些实施例中,推送单元包括:第一推送子单元,被配置成将至少一个网页内容片段的主题推送给用户。

[0014] 在一些实施例中,推送单元包括:第一生成子单元,被配置成对于至少一个网页内容片段中的网页内容片段,将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字,对匹配的文字设置预设文字效果,生成设置文字效果的网页;第二推送子单元,被配置成将设置文字效果的网页推送给用户。

[0015] 在一些实施例中,推送单元包括:切词子单元,被配置成对至少一个网页内容片段的主题进行切词,基于切词结果生成网页内容对应的词向量;预测子单元,被配置成将网页内容对应的词向量输入至预先训练的意图预测模型,得到网页内容对应的搜索意图,其中,意图预测模型用于预测搜索意图;提取子单元,被配置成从用户的用户画像中提取兴趣标签集合;匹配子单元,被配置成将网页内容对应的搜索意图在兴趣标签集合中进行匹配,获取匹配的兴趣标签;第二生成子单元,被配置成将网页内容中的文字与匹配的兴趣标签进行匹配,获取匹配的文字,对匹配的文字设置预设文字效果,生成设置文字效果的网页;第三推送子单元,被配置成将设置文字效果的网页推送给用户。

[0016] 在一些实施例中,意图预测模型通过如下步骤训练得到:获取训练样本集合,其中,训练样本包括样本搜索数据和样本搜索数据对应的样本搜索意图;基于训练样本集合,对多层神经网络进行训练,得到意图预测模型。

[0017] 第三方面,本申请实施例提供了一种服务器,该服务器包括:一个或多个处理器;存储装置,其上存储有一个或多个程序;当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

[0018] 第四方面,本申请实施例提供了一种计算机可读介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

[0019] 本申请实施例提供的用于推送信息的方法和装置,在用户选择历史搜索数据时,获取历史搜索数据对应的网页内容;然后将网页内容划分成至少一个网页内容片段,并提取至少一个网页内容片段的主题;最后基于至少一个网页内容片段的主题,向用户推送信息。在用户再次浏览网页时,实现了快速地向用户推送网页的核心信息。

附图说明

[0020] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

- [0021] 图1是本申请可以应用于其中的示例性系统架构；
- [0022] 图2是根据本申请的用于推送信息的方法的一个实施例的流程图；
- [0023] 图3是历史搜索数据集合的示意图；
- [0024] 图4是根据本申请的用于推送信息的方法的又一个实施例的流程图；
- [0025] 图5是根据本申请的用于推送信息的方法的再一个实施例的流程图；
- [0026] 图6是根据本申请的用于推送信息的装置的一个实施例的结构示意图；
- [0027] 图7是适于用来实现本申请实施例的服务器的计算机系统的结构示意图。

具体实施方式

[0028] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

[0029] 需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0030] 图1示出了可以应用本申请的用于推送信息的方法或用于推送信息的装置的实施例的示例性系统架构100。

[0031] 如图1所示，系统架构100中可以包括终端设备101、102、103，网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

[0032] 用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。终端设备101、102、103上可以安装有各种客户端应用，例如网页浏览器应用等。

[0033] 终端设备101、102、103可以是硬件，也可以是软件。当终端设备101、102、103为硬件时，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。当终端设备101、102、103为软件时，可以安装在上述电子设备中。其可以实现成多个软件或软件模块，也可以实现成单个软件或软件模块。在此不做具体限定。

[0034] 服务器105可以提供各种服务，例如服务器105可以对从终端设备101、102、103获取到的历史搜索数据等数据进行分析等处理，并将处理结果（例如推送信息）反馈给终端设备101、102、103。

[0035] 需要说明的是，服务器105可以是硬件，也可以是软件。当服务器105为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器105为软件时，可以实现成多个软件或软件模块（例如用来提供分布式服务），也可以实现成单个软件或软件模块。在此不做具体限定。

[0036] 需要说明的是，本申请实施例所提供的用于推送信息的方法一般由服务器105执行，相应地，用于推送信息的装置一般设置于服务器105中。

[0037] 应该理解，图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器。

[0038] 继续参考图2，其示出了根据本申请的用于推送信息的方法的一个实施例的流程200。该用于推送信息的方法，包括以下步骤：

[0039] 步骤201, 响应于用户选择历史搜索数据, 获取历史搜索数据对应的网页内容。

[0040] 在本实施例中, 当用户选择历史搜索数据时, 用于推送信息的方法的执行主体 (例如如图1所示的服务器105) 可以获取历史搜索数据对应的网页的内容。其中, 历史搜索数据可以包括用户之前输入的搜索词和用户之前浏览过的网页的主题。

[0041] 实践中, 当用户需要浏览网页时, 可以打开终端设备 (例如如图1所示的终端设备101、102、103) 上安装的网页浏览器应用, 并在网页浏览器应用的搜索框中输入搜索词。随后, 用户点击网页浏览器应用中的搜索按钮, 向上述执行主体发起搜索请求。当上述执行主体接收到搜索请求时, 可以利用用户输入的搜索词在搜索引擎中进行搜索, 以获取与搜索词关联的网页集合, 并将与搜索词关联的网页集合对应的主题集合发送给用户, 以供用户的终端设备进行展示。当用户点击主题集合中的主题时, 可以向上述执行主体发起浏览请求。当上述执行主体接收到浏览请求时, 可以从与搜索词关联的网页集合中获取用户点击的主题对应的网页, 并发送给用户, 以供用户进行浏览。由此可知, 在搜索过程中, 上述执行主体不仅可以获取用户输入的搜索词, 还可以获取用户浏览过的网页的主题。这样, 上述执行主体就可以基于用户在预设时间段内 (例如最近一个小时内) 输入的搜索词和浏览过的网页的主题聚合成历史搜索数据集合。其中, 历史搜索数据集合中的一条历史搜索数据可以包括用户输入的一个搜索词和用户浏览过的一个网页。

[0042] 作为示例, 若用户在最近一个小时内网页浏览器应用中输入过搜索词“婴儿床什么牌子好”, 并浏览过主题为“婴儿床十大品牌排名”、“婴儿床什么牌子好”的网页, 以及输入过搜索词“某某品牌的婴儿床怎么样”, 并浏览过主题为“某某品牌的婴儿床怎么样”和“某某品牌的婴儿床价格”的网页, 那么上述执行主体聚合出历史搜索数据集合, 具体如图3所示。网页浏览器应用中可以设置“近期关注”按钮, 当用户需要再次浏览这些网页时, 可以点击“近期关注”按钮。此时, 上述执行主体可以将历史搜索数据集合发送至用户的终端设备进行展示。其中, 历史搜索数据集合中的网页的主题可以与网页的网址相关联。当用户直接点击历史搜索数据集合中的某个网页的主题时, 可以直接跳转展示该网页, 以供用户再次浏览该网页的内容。历史搜索数据集合中的每个网页的主题的附近 (例如网页的主题的后面) 可以设置“加入对比”按钮, 当用户点击某个网页的主题的附近的“加入对比”按钮时, 就选中了该网页的主题。历史搜索数据集合的附近 (例如历史搜索数据集合的后面) 可以设置“开始对比”按钮。当用户点击“开始对比”按钮时, 就可以将用户选中的网页的主题发送至上述执行主体。例如, 用户点击了主题“婴儿床十大品牌排名”和“某某品牌的婴儿床怎么样”附近的“加入对比”按钮, 并点击了“开始对比”按钮。此时, 上述执行主体可以根据主题“婴儿床十大品牌排名”和“某某品牌的婴儿床怎么样”查找出相应的网页, 并获取网页的内容。

[0043] 步骤202, 将网页内容划分成至少一个网页内容片段, 提取至少一个网页内容片段的主题。

[0044] 在本实施例中, 上述执行主体可以将网页内容划分成至少一个网页内容片段。对于至少一个网页内容片段中的每个网页内容片段, 上述执行主体可以提取该网页内容片段的主题。例如, 若网页内容包括 N (N 为正整数) 个段落, 上述执行主体可以将网页内容划分成 N 个网页内容片段。其中, 一个网页内容片段对应网页内容的一个段落。又例如, 上述执行主体可以将网页内容随机划分为多个网页内容片段。其中, 所划分出的网页内容片段的数目

可以是默认设置的数目,也可以是用户根据需求自主设置的数目。

[0045] 在本实施例的一些可选的实现方式中,对于至少一个网页内容片段中的网页内容片段,上述执行主体可以将该网页内容片段输入至预先训练的文档主题生成模型,得到该网页内容片段的主题。其中,文档主题生成模型可以用于生成文档的主题的概率分布。具体地,对于至少一个网页内容片段中的每个网页内容片段,上述执行主体可以将该网页内容片段输入至文档主题生成模型,输出该网页内容片段的一些主题的概率分布,根据该网页内容片段的一些主题的概率分布,可以得到该网页内容片段的主题。其中,网页内容片段的主题的数目可以是不小于零的整数,即网页内容片段可以没有主题,也可以仅有一个主题,还可以有多个主题。

[0046] 作为示例,文档主题生成模型可以是LDA(Latent Dirichlet Allocation)。其中,LDA也称为三层贝叶斯概率模型,包含词、主题和文档三层结构。LDA是一种非监督机器学习技术,可以用来识别大规模文档集(document collection)或语料库(corpus)中潜藏的主题。它采用了词袋(bag of words)的方法,这种方法将每一篇文档视为一个词频向量,从而将文本信息转化为了易于建模的数字信息。每一篇文档代表了一些主题所构成的一个概率分布,而每一个主题又代表了很多单词所构成的一个概率分布。

[0047] 步骤203,基于至少一个网页内容片段的主题,向用户推送信息。

[0048] 在本实施例中,上述执行主体可以基于至少一个网页内容片段的主题,向用户推送信息。例如,上述执行主体可以将至少一个网页内容片段的主题推送给用户。又例如,上述执行主体可以基于至少一个网页内容片段的主题对网页内容进行编辑,生成新网页,并将新网页推送给用户。

[0049] 需要说明的是,当用户需要重新浏览多个网页时,可以选择多条历史搜索数据,对于多条历史搜索数据中的每条历史搜索数据,均执行步骤201-203。这样,用户就可以接收到多条历史搜索数据对应的多个推送信息。这样,用户就可以根据多条推送信息对多个网页的核心内容进行比对。

[0050] 本申请实施例提供的用于推送信息的方法,在用户选择历史搜索数据时,获取历史搜索数据对应的网页内容;然后将网页内容划分成至少一个网页内容片段,并提取至少一个网页内容片段的主题;最后基于至少一个网页内容片段的主题,向用户推送信息。在用户再次浏览历史网页时,实现了快速地向用户推送网页的核心信息。

[0051] 进一步参考图4,其示出了根据本申请的用于推送信息的方法的又一个实施例的流程400。该用于推送信息的方法,包括以下步骤:

[0052] 步骤401,响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容。

[0053] 步骤402,将网页内容划分成至少一个网页内容片段,提取至少一个网页内容片段的主题。

[0054] 在本实施例中,步骤401-402的具体操作与图2所示的实施例中步骤201-202的操作基本相同,在此不再赘述。

[0055] 步骤403,将至少一个网页内容片段的主题推送给用户。

[0056] 在本实施例中,用于推送信息的方法的执行主体(例如图1所示的服务器105)可以将至少一个网页内容片段的主题推送给用户,以供用户的终端设备(例如图1所示的终端设备101、102、103)进行显示。

[0057] 步骤404,对于至少一个网页内容片段中的网页内容片段,将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字。

[0058] 在本实施例中,对于至少一个网页内容片段中的每个网页内容片段,上述执行主体可以将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字。作为示例,上述执行主体可以计算该网页内容片段中的每条语句与该网页内容片段的主题的相似度,若相似度大于预设相似度阈值,则将该语句中的文字作为匹配的文字。这里,对于该网页内容片段中的每条语句,上述执行主体可以对该语句进行切词,生成该语句的关键词集合。同样,上述执行主体可以对该网页内容片段的主题进行切词,生成该网页内容片段的主题的关键词集合。而后,上述执行主体可以计算该语句的关键词集合与该网页内容片段的主题的关键词集合的相似度。例如,可以将该语句的关键词集合与该网页内容片段的主题的关键词集合中相同关键词的数目与该语句的关键词集合中的关键词的数目的比值作为该语句与该网页内容片段的主题的相似度。

[0059] 步骤405,对匹配的文字设置预设文字效果,生成设置文字效果的网页。

[0060] 在本实施例中,上述执行主体可以对匹配的文字设置预设文字效果,生成设置文字效果的网页。这里,对匹配的文字设置文字效果可以使匹配的文字与不匹配的文字相区别,以使用户在浏览网页时,快速地获取匹配的文字。例如,可以将匹配的文字的颜色设置成黄色,以便于用户在浏览设置文字效果的网页时,可以第一时间注意到设置成黄色的文字。

[0061] 步骤406,将设置文字效果的网页推送给用户。

[0062] 在本实施例中,上述执行主体可以将设置文字效果的网页推送给用户。通常,终端设备上会先显示至少一个网页内容片段的主题。当用户点击一个网页内容片段的主题时,可以展示设置文字效果的网页中的该网页内容片段。其中,展示的该网页内容片段中与该网页内容片段的主题匹配的文字被设置的文字效果,以便于用户在查看该网页内容片段时,可以第一时间注意到设置文字效果的文字。

[0063] 从图4中可以看出,与图2对应的实施例相比,本实施例中的用于推送信息的方法的流程400突出了向用户推送信息的步骤。由此,本实施例描述的方案分层次展现推送信息。首先展示至少一个网页内容片段的主题。当用户点击一个网页内容片段的主题时,再展示设置文字效果的网页中的该网页内容片段。由于展示的该网页内容片段中与该网页内容片段的主题匹配的文字被设置的文字效果,以便于用户再次浏览该网页内容片段时,可以第一时间注意到设置文字效果的文字。不仅可以使用户快速地获取网页的核心信息,还可以使用户更全面地获取网页的核心信息。

[0064] 进一步参考图5,其示出了根据本申请的用于推送信息的方法的再一个实施例的流程500。该用于推送信息的方法,包括以下步骤:

[0065] 步骤501,响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容。

[0066] 步骤502,将网页内容划分成至少一个网页内容片段,提取至少一个网页内容片段的主题。

[0067] 在本实施例中,步骤501-502的具体操作与图2所示的实施例中步骤201-202的操作基本相同,在此不再赘述。

[0068] 步骤503,对至少一个网页内容片段的主题进行切词,基于切词结果生成网页内容

对应的词向量。

[0069] 在本实施例中,对于至少一个网页内容片段中的每个网页内容片段,用于推送信息的方法的执行主体(例如图1所示的服务器105)可以首先对该网页内容片段的主题进行切词,得到该网页内容片段的主题的关键词集合;然后基于该网页内容片段的主题的关键词集合生成词向量中的一维。因此,网页内容对应的词向量的维数与网页内容划分成的网页内容片段的个数相同。

[0070] 步骤504,将网页内容对应的词向量输入至预先训练的意图预测模型,得到网页内容对应的搜索意图。

[0071] 在本实施例中,上述执行主体可以将网页内容对应的词向量输入至预先训练的意图预测模型,从而输出网页内容对应的搜索意图。其中,意图预测模型可以用于预测搜索意图,表征词向量与搜索意图之间的对应关系。其中,搜索意图可以用于表达用户的搜索动机,即在发起搜索时,用户内心真实想要获取的信息。

[0072] 在本实施例的一些可选的实现方式中,意图预测模型可以是本领域技术人员对大量词向量以及对应的搜索意图进行统计分析,而得到的存储有多个词向量与对应的搜索意图的对应关系表。此时,上述执行主体可以计算网页内容对应的词向量与对应关系表中的各个词向量之间的相似度,并基于相似度计算结果,从该对应关系表中得到网页内容对应的搜索意图。例如,首先确定出与网页内容对应的词向量相似度最高的词向量,然后从该对应关系表中查找出与网页内容对应的词向量相似度最高的词向量对应的搜索意图,作为该网页内容对应的搜索意图。

[0073] 在本实施例的一些可选的实现方式中,意图预测模型可以是利用各种机器学习方法和训练样本对现有的机器学习模型(例如各种人工神经网络等)进行有监督训练而得到的。这里,目标检测模型可以是MLP(Multi-Layer Perceptron,多层神经网络),通过如下步骤训练得到:

[0074] 首先,获取训练样本集合。

[0075] 其中,训练样本可以包括样本搜索数据和样本搜索数据对应的样本搜索意图。样本搜索数据可以包括样本搜索词和样本网页的主题。这里,本领域技术人员可以对样本搜索数据进行分析,从而得到样本搜索数据对应的样本搜索意图。例如,本领域技术人员可以对大量输入样本搜索词和浏览样本网页的用户进行问卷调查,得到这些用户的搜索意图。随后,对这些用户的搜索意图进行归纳总结,得到样本搜索意图。

[0076] 其次,基于训练样本集合,对多层神经网络进行训练,得到意图预测模型。

[0077] 这里,对于训练样本集合中的训练样本,上述执行主体可以首先对该训练样本中的样本搜索数据进行切词,基于切词结果生成该样本搜索数据对应的词向量;然后将该样本搜索数据对应的词向量作为输入,将样本搜索数据对应的样本搜索意图作为输出,对多层神经网络进行训练,从而得到意图预测模型。其中,多层神经网络可以是未经训练或未训练完成的多层神经网络。这里,对于未经训练的多层神经网络,其各个参数(例如,权值参数和偏置参数)用一些不同的小随机数进行初始化。“小随机数”用来保证网络不会因权值过大而进入饱和状态,从而导致训练失败,“不同”用来保证网络可以正常地学习。对于未训练完成的多层神经网络,其各个参数可以是被调整后参数,但是该多层神经网络的预测效果通常尚未满足预先设置的约束条件。

[0078] 步骤505,从用户的用户画像中提取兴趣标签集合。

[0079] 在本实施例中,上述执行主体可以从用户的用户画像中提取至少部分兴趣标签,生成兴趣标签集合。例如,上述执行主体可以按照权重从大到小的顺序对用户的用户画像中的兴趣标签进行排序,选取权重排序在前预设位(例如前10位)的兴趣标签,生成兴趣标签集合。

[0080] 实践中,用户画像可以是根据预先积累的用户数据所建立的包括至少一种类别的描述性标签的用户信息。描述性标签可以用于描述用户的性别、年龄、婚姻状况、职业、收入状况、教育水平、兴趣等等。其中,用于描述用户的兴趣的描述性标签就是兴趣标签。不同的兴趣标签可以设置不同的权重,通常用户对某个事物兴趣越高,其对应的兴趣标签的权重越大,反之,权重越小。作为示例,一个用户画像所包含的用户信息可以是“性别:男,年龄:30-35岁,婚姻状况:已婚,职业:银行职员,收入状况:大于8000,教育水平:大学毕业,兴趣:体育、财经、科技”。该用户的兴趣标签可以包括:体育、财经、科技。

[0081] 步骤506,将网页内容对应的搜索意图在兴趣标签集合中进行匹配,获取匹配的兴趣标签。

[0082] 在本实施例中,上述执行主体可以计算网页内容对应的搜索意图与兴趣标签集合中的每个兴趣标签的相似度,基于相似度确定匹配的兴趣标签。例如,将相似度最高的兴趣标签作为匹配的兴趣标签。

[0083] 步骤507,将网页内容中的文字与匹配的兴趣标签进行匹配,获取匹配的文字。

[0084] 在本实施例中,对于至少一个网页内容片段中的每个网页内容片段,上述执行主体可以将该网页内容片段中的文字与匹配的兴趣标签进行匹配,获取匹配的文字。作为示例,上述执行主体可以计算该网页内容片段中的每条语句与匹配的兴趣标签的相似度,若相似度大于预设相似度阈值,则将该语句中的文字作为匹配的文字。

[0085] 步骤508,对匹配的文字设置预设文字效果,生成设置文字效果的网页。

[0086] 步骤509,将设置文字效果的网页推送给用户。

[0087] 在本实施例中,步骤508-509的具体操作与图4所示的实施例中步骤405-406的操作基本相同,在此不再赘述。

[0088] 从图5中可以看出,与图2对应的实施例相比,本实施例中的用于推送信息的方法的流程500突出了向用户推送信息的步骤。由此,本实施例描述的方案展示设置文字效果的网页。由于网页内容中与匹配的兴趣标签匹配的文字被设置的文字效果,以便于用户再次浏览网页内容时,可以第一时间注意到设置文字效果的文字。不仅可以使用户快速地获取网页的核心信息,还可以使用户更准确地获取网页的核心信息。

[0089] 进一步参考图6,作为对上述各图所示方法的实现,本申请提供了一种用于推送信息的装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0090] 如图6所示,本实施例的用于推送信息的装置600可以包括:获取单元601、提取单元602和推送单元603。其中,获取单元601,被配置成响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容;提取单元602,被配置成将网页内容划分成至少一个网页内容片段,提取至少一个网页内容片段的主题;推送单元603,被配置成基于至少一个网页内容片段的主题,向用户推送信息。

[0091] 在本实施例中,用于推送信息的装置600中:获取单元601、提取单元602和推送单元603的具体处理及其所带来的技术效果可分别参考图2对应实施例中的步骤201、步骤202和步骤203的相关说明,在此不再赘述。

[0092] 在本实施例的一些可选的实现方式中,提取单元602可以进一步被配置成:对于至少一个网页内容片段中的网页内容片段,将该网页内容片段输入至预先训练的文档主题生成模型,得到该网页内容片段的主题。

[0093] 在本实施例的一些可选的实现方式中,推送单元603可以包括:第一推送子单元(图中未示出),被配置成将至少一个网页内容片段的主题推送给用户。

[0094] 在本实施例的一些可选的实现方式中,推送单元603可以包括:第一生成子单元(图中未示出),被配置成对于至少一个网页内容片段中的网页内容片段,将该网页内容片段中的文字与该网页内容片段的主题进行匹配,获取匹配的文字,对匹配的文字设置预设文字效果,生成设置文字效果的网页;第二推送子单元(图中未示出),被配置成将设置文字效果的网页推送给用户。

[0095] 在本实施例的一些可选的实现方式中,推送单元603可以包括:切词子单元(图中未示出),被配置成对至少一个网页内容片段的主题进行切词,基于切词结果生成网页内容对应的词向量;预测子单元(图中未示出),被配置成将网页内容对应的词向量输入至预先训练的意图预测模型,得到网页内容对应的搜索意图,其中,意图预测模型用于预测搜索意图;提取子单元(图中未示出),被配置成从用户的用户画像中提取兴趣标签集合;匹配子单元(图中未示出),被配置成将网页内容对应的搜索意图在兴趣标签集合中进行匹配,获取匹配的兴趣标签;第二生成子单元(图中未示出),被配置成将网页内容中的文字与匹配的兴趣标签进行匹配,获取匹配的文字,对匹配的文字设置预设文字效果,生成设置文字效果的网页;第三推送子单元(图中未示出),被配置成将设置文字效果的网页推送给用户。

[0096] 在本实施例的一些可选的实现方式中,意图预测模型通过如下步骤训练得到:获取训练样本集合,其中,训练样本包括样本搜索数据和样本搜索数据对应的样本搜索意图;基于训练样本集合,对多层神经网络进行训练,得到意图预测模型。

[0097] 下面参考图7,其示出了适于用来实现本申请实施例的服务器(例如图1所示的服务器105)的计算机系统700的结构示意图。图7示出的服务器仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0098] 如图7所示,计算机系统700包括中央处理单元(CPU)701,其可以根据存储在只读存储器(ROM)702中的程序或者从存储部分708加载到随机访问存储器(RAM)703中的程序而执行各种适当的动作和处理。在RAM 703中,还存储有系统700操作所需的各种程序和数据。CPU 701、ROM 702以及RAM 703通过总线704彼此相连。输入/输出(I/O)接口705也连接至总线704。

[0099] 以下部件连接至I/O接口705:包括键盘、鼠标等的输入部分706;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707;包括硬盘等的存储部分708;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709。通信部分709经由诸如因特网的网络执行通信处理。驱动器710也根据需要连接至I/O接口705。可拆卸介质711,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710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708。

[0100] 特别地,根据本公开的实施例,上文参考流程图描述的过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709从网络上被下载和安装,和/或从可拆卸介质711被安装。在该计算机程序被中央处理单元(CPU)701执行时,执行本申请的方法中限定的上述功能。需要说明的是,本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读介质或者是上述两者的任意组合。计算机可读介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、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)、光纤、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。在本申请中,计算机可读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。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0101] 可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码,所述程序设计语言包括面向目标的设计语言——诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言——诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0102] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、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,该模块、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,框图和/或流程图中的每个方框、以及框图和/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。

[0103] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器包

括获取单元、提取单元和推送单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,获取单元还可以被描述为“响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容的单元”。

[0104] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的服务器中所包含的;也可以是单独存在,而未装配入该服务器中。上述计算机可读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该服务器执行时,使得该服务器:响应于用户选择历史搜索数据,获取历史搜索数据对应的网页内容;将网页内容划分成至少一个网页内容片段,提取至少一个网页内容片段的主题;基于至少一个网页内容片段的主题,向用户推送信息。

[0105] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本申请中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

100

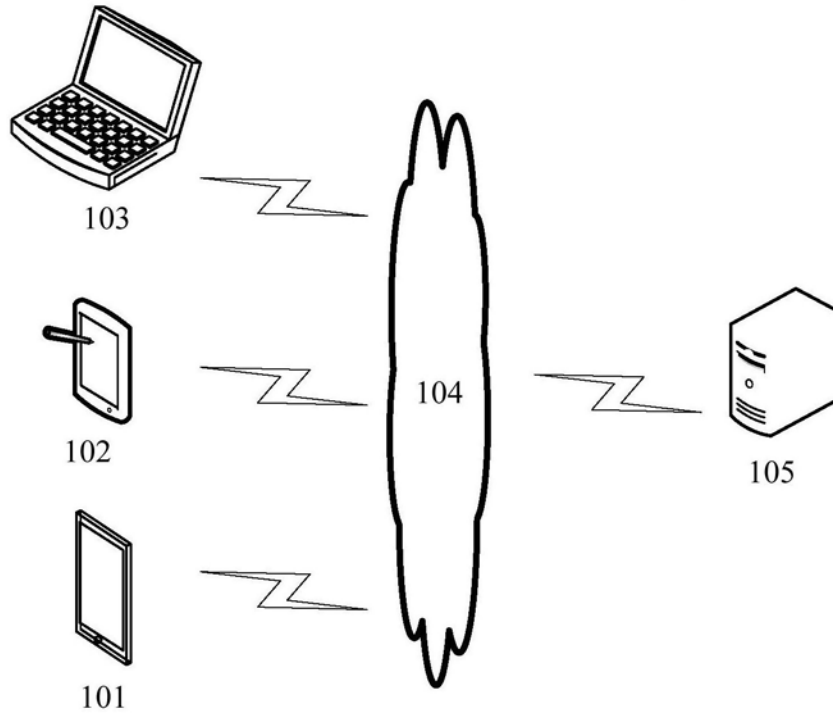


图1

200

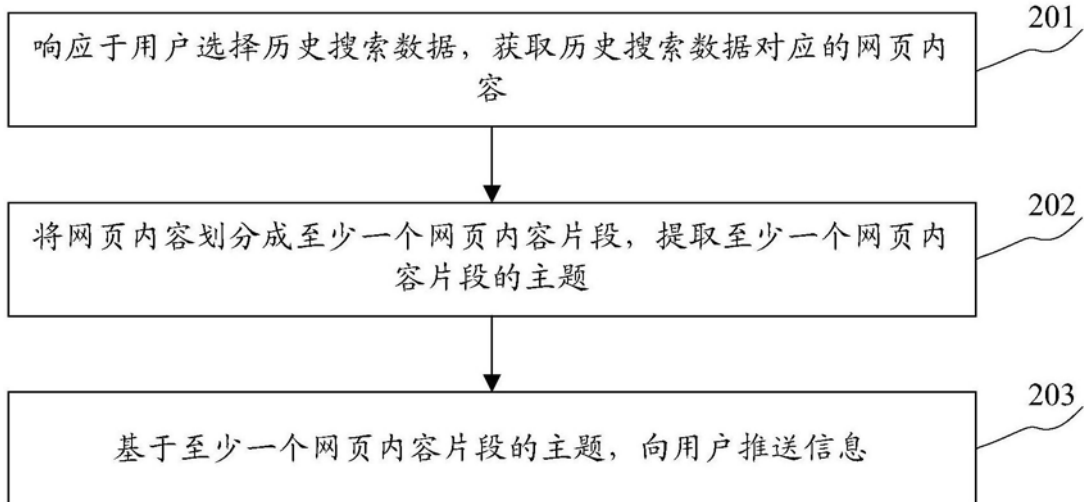


图2

您输入过的搜索词:	您浏览过的网页的主题:
婴儿床什么牌子好	婴儿床十大品牌排名
	婴儿床什么牌子好
某某品牌的婴儿床怎么样	某某品牌的婴儿床怎么样
	某某品牌的婴儿床价格

图3

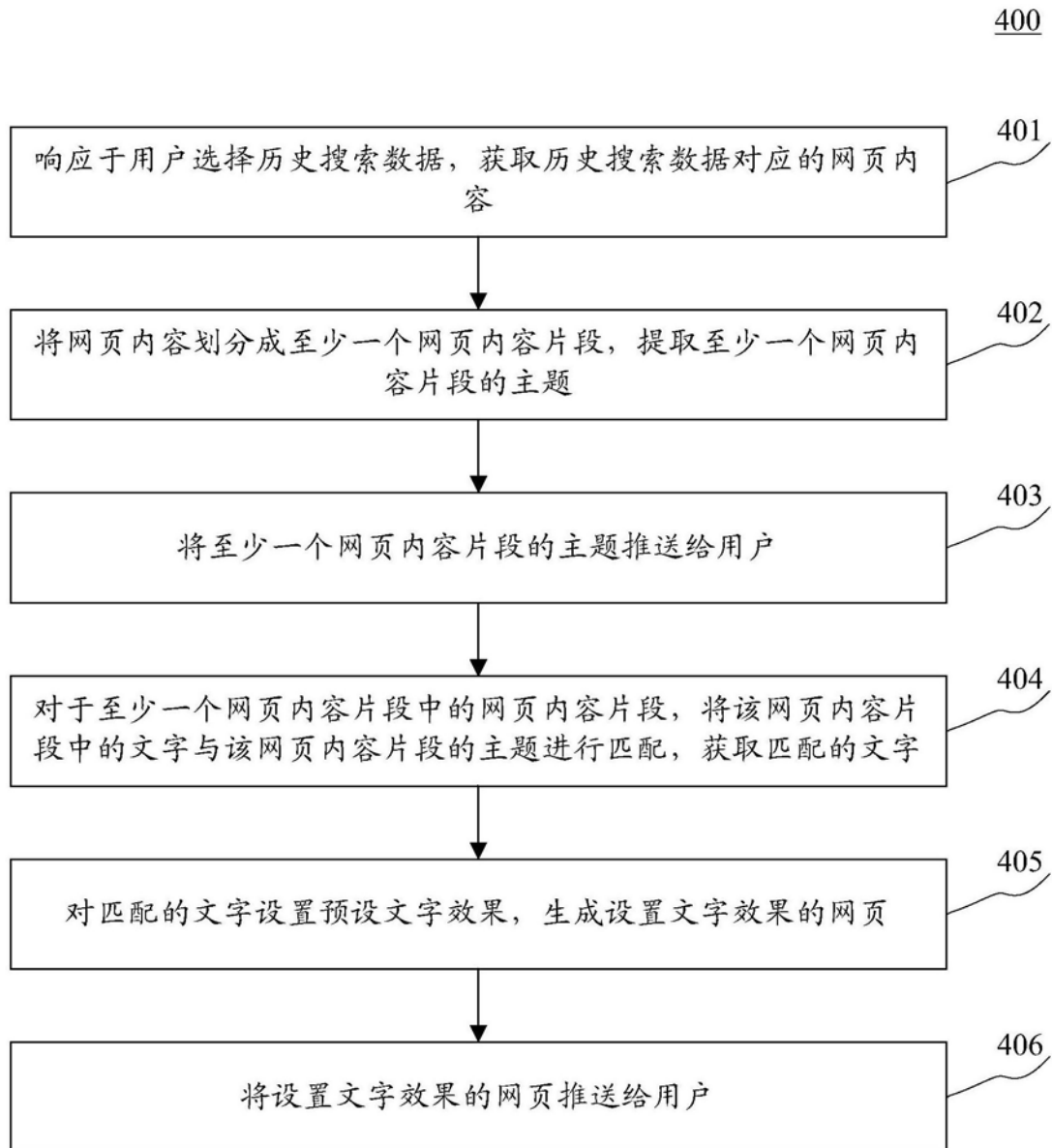


图4

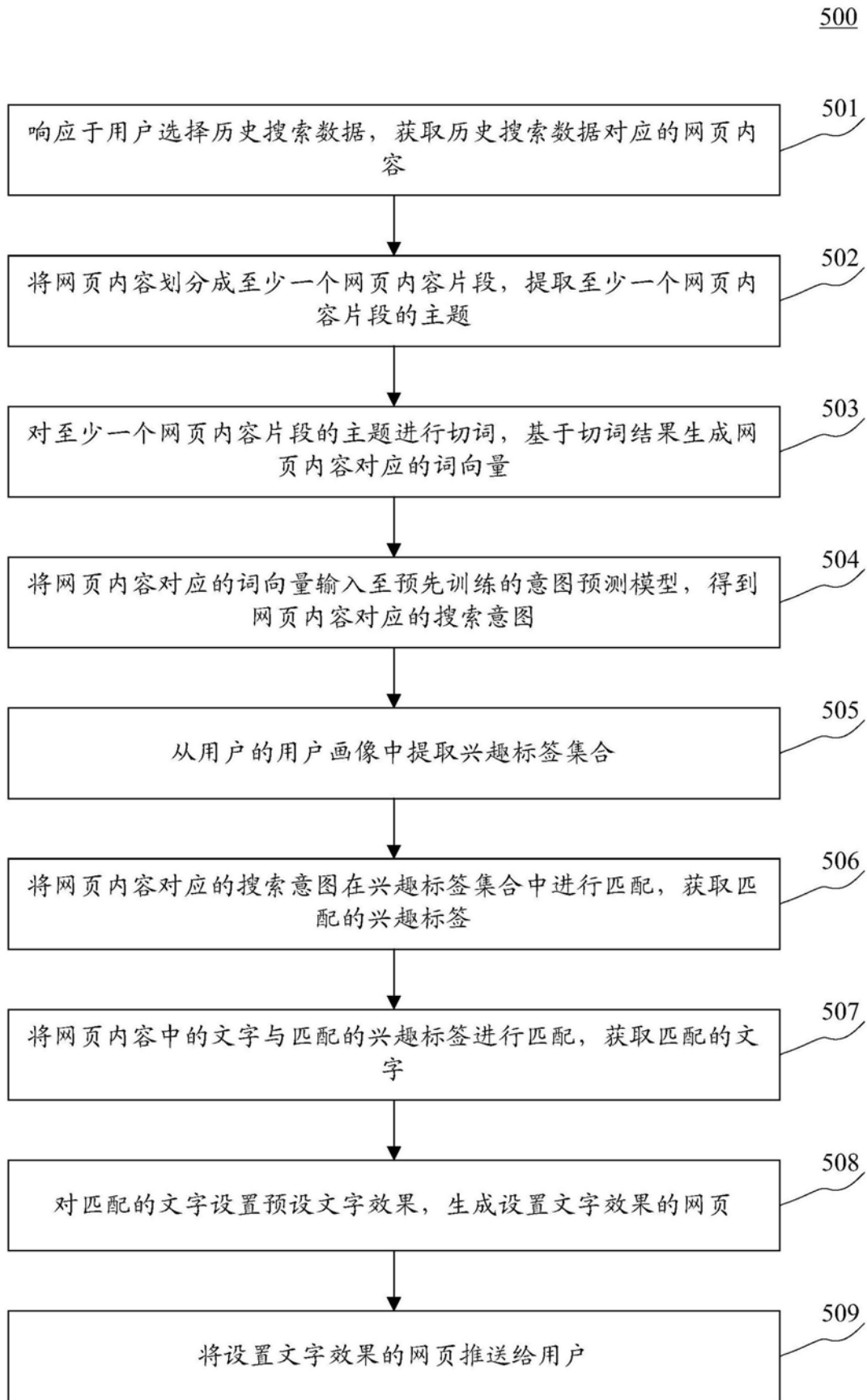


图5

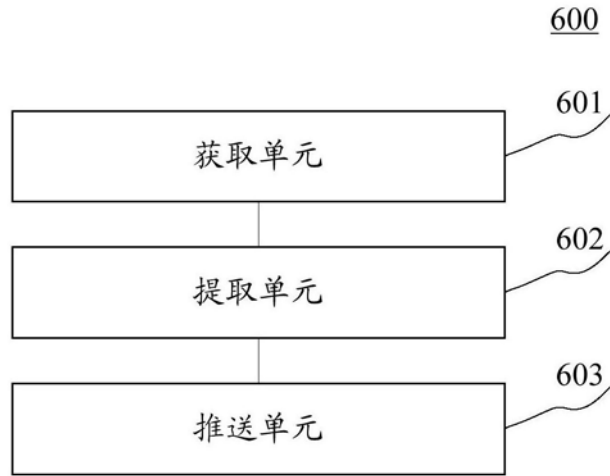


图6

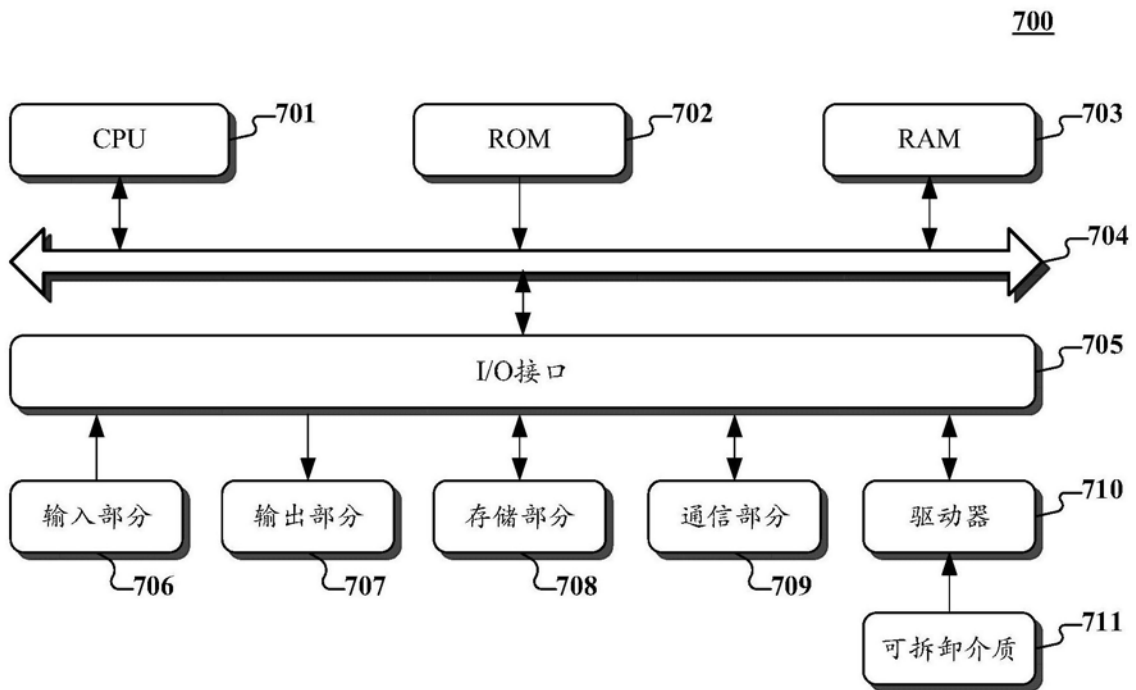


图7