



(12) 发明专利

(10) 授权公告号 CN 110232156 B

(45) 授权公告日 2022.08.19

(21) 申请号 201910473094.7

(22) 申请日 2019.05.31

(65) 同一申请的已公布的文献号  
申请公布号 CN 110232156 A

(43) 申请公布日 2019.09.13

(73) 专利权人 腾讯科技(深圳)有限公司  
地址 518000 广东省深圳市南山区高新区  
科技中一路腾讯大厦35层

(72) 发明人 王卓然 亓超 马宇驰 陈华荣  
温泉 范彦革 梁伟 岳媛媛  
刁德纯 曹圣明 李宇舰 王东亮  
赵巍 林梓悦

(74) 专利代理机构 深圳市深佳知识产权代理事  
务所(普通合伙) 44285  
专利代理师 王仲凯

(51) Int.Cl.

G06F 16/9535 (2019.01)

G06F 40/289 (2020.01)

G06F 40/295 (2020.01)

G06F 40/30 (2020.01)

(56) 对比文件

CN 109800352 A, 2019.05.24

CN 109285030 A, 2019.01.29

US 2010/0076991 A1, 2010.03.25

审查员 张杨悦

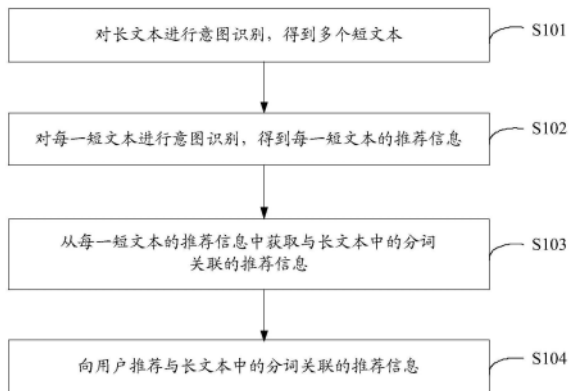
权利要求书2页 说明书9页 附图3页

(54) 发明名称

基于长文本的信息推荐方法及装置

(57) 摘要

本发明实施例公开了一种基于长文本的信息推荐方法,包括:对长文本进行意图识别,得到多个短文本,所述短文本为基于所述长文本中的分词扩展出的文本;对每一短文本进行意图识别,得到所述每一短文本的推荐信息;从所述每一短文本的推荐信息中获取与所述长文本中的分词关联的推荐信息;向用户推荐与所述长文本中的分词关联的推荐信息。由于向用户推荐的推荐信息是根据长文本中的分词从每一短文本的推荐信息中获取的,因此,能够减少推荐信息的数量,以使用户查找,并且,由于每一短文本是通过长文本进行意图识别得到的,因此,还能够使推荐的信息更加精准。



1. 一种基于长文本的信息推荐方法,其特征在于,包括:

获取所述长文本中的多个分词;

采用排列组合中组合的方式,将所述长文本中的至少两个分词进行两两合并,和/或三三合并,得到多个合并后的分词;

计算所述多个合并后的分词的权重值,删除权重值小于预设阈值的合并后的分词,保留权重值大于或等于预设阈值的合并后的分词,将权重值大于或等于预设阈值的合并后的分词作为长文本对应的多个短文本;或者对保留的合并后的分词进行扩展,扩展出短文本,将扩展出的短文本作为长文本对应的多个短文本;所述合并后的分词的权重值大小用于标识合并后的分词对应的推荐信息的多少;

对每一短文本进行分词处理,得到分词结果;基于分词结果进行命名实体识别,识别出短文本的意图;基于所述短文本的意图将短文本输入到搜索引擎中进行搜索,将搜索结果作为推荐信息,得到所述每一短文本的推荐信息;

确定所述每一短文本的推荐信息的数量;

若所述每一短文本的推荐信息的数量大于或等于第一预设数量,则基于所述长文本中的分词,对所述每一短文本的推荐信息进行筛选,获得与所述长文本中的分词关联的推荐信息;

若所述每一短文本的推荐信息的数量小于第一预设数量,则确定所述多个短文本的数量;

当所述多个短文本的数量大于或等于第二预设数量时,则基于所述长文本中的分词,对所述多个短文本的推荐信息进行筛选,获得与所述长文本中的分词关联的推荐信息;长文本中的分词与短文本中的分词是不同的分词;

向用户推荐与所述长文本中的分词关联的推荐信息。

2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

获所述长文本中的多个分词;

对所述长文本中的每一分词进行意图识别,得到所述每一分词对应的短文本,并将所述每一分词对应的短文本作为所述多个短文本。

3. 一种基于长文本的信息推荐装置,其特征在于,包括:

获取模块,用于获取所述长文本中的多个分词;采用排列组合中组合的方式,将所述长文本中的至少两个分词进行两两合并,和/或三三合并,得到多个合并后的分词;计算所述多个合并后的分词的权重值,删除权重值小于预设阈值的合并后的分词,保留权重值大于或等于预设阈值的合并后的分词,将权重值大于或等于预设阈值的合并后的分词作为长文本对应的多个短文本;或者对保留的合并后的分词进行扩展,扩展出短文本,将扩展出的短文本作为长文本对应的多个短文本;所述合并后的分词的权重值大小用于标识合并后的分词对应的推荐信息的多少;

识别模块,用于对每一短文本进行分词处理,得到分词结果;基于分词结果进行命名实体识别,识别出短文本的意图;基于所述短文本的意图将短文本输入到搜索引擎中进行搜索,将搜索结果作为推荐信息,得到所述每一短文本的推荐信息;

确定模块,用于确定所述每一短文本的推荐信息的数量;若所述每一短文本的推荐信息的数量大于或等于第一预设数量,则基于所述长文本中的分词,对所述每一短文本的推

荐信息进行筛选,获得与所述长文本中的分词关联的推荐信息;若所述每一短文本的推荐信息的数量小于第一预设数量,则确定所述多个短文本的数量;当所述多个短文本的数量大于或等于第二预设数量时,则基于所述长文本中的分词,对所述多个短文本的推荐信息进行筛选,获得与所述长文本中的分词关联的推荐信息;长文本中的分词与短文本中的分词是不同的分词;

推荐模块,用于向用户推荐与所述长文本中的分词关联的推荐信息。

4. 根据权利要求3所述的装置,其特征在于,所述装置还包括:

第一获取单元,用于获取所述长文本中的多个分词;

第一识别单元,用于对所述长文本中的每一分词进行意图识别,得到所述每一分词对应的短文本,并将所述每一分词对应的短文本作为所述多个短文本。

5. 一种电子设备,其特征在于,包括:

至少一个处理器;

以及与所述处理器连接的至少一个存储器、总线;

其中,所述处理器、存储器通过所述总线完成相互间的通信;所述处理器用于调用所述存储器中的程序指令,以执行如权利要求1至2中任一项所述的方法。

6. 一种计算机可读存储介质,其特征在于,所述存储介质包括存储的程序,其中,在所述程序运行时控制所述存储介质所在设备执行如权利要求1至2中任一项所述的方法。

## 基于长文本的信息推荐方法及装置

### 技术领域

[0001] 本发明实施例涉及信息处理技术领域,特别是涉及一种基于长文本的信息推荐方法及装置。

### 背景技术

[0002] 长文本,是指其中存在有多个分词(即具有实际意义的词语)的文本。在现有技术中,如果需要根据长文本向用户推荐信息,一般的,是对长文本进行分词,按照分词获取对应的热点内容或者是与分析相关的用户喜好的推荐内容,长文本分词较多,推荐内容也较多,用户想要获取感兴趣的内容需要在大量信息中搜索,非常不便。

### 发明内容

[0003] 有鉴于此,本发明实施例提供了一种基于长文本的信息推荐方法及装置,能够减少向用户推荐的信息的数量,并且使得推荐的信息更加精准。

[0004] 为了解决上述问题,本发明实施例主要提供如下技术方案:

[0005] 第一方面,本发明实施例提供了一种基于长文本的信息推荐方法,该方法包括:对长文本进行意图识别,得到多个短文本,所述短文本为基于所述长文本中的分词扩展出的文本;对每一短文本进行意图识别,得到所述每一短文本的推荐信息;从所述每一短文本的推荐信息中获取与所述长文本中的分词关联的推荐信息;向用户推荐与所述长文本中的分词关联的推荐信息。

[0006] 第二方面,本发明实施例还提供一种基于长文本的信息推荐装置,该装置包括:获取模块,用于对长文本进行意图识别,得到多个短文本,所述短文本为基于所述长文本中的分词扩展出的文本;识别模块,用于对每一短文本进行意图识别,得到所述每一短文本的推荐信息;确定模块,用于从所述每一短文本的推荐信息中获取与所述长文本中的分词关联的推荐信息;推荐模块,用于向用户推荐与所述长文本中的分词关联的推荐信息。

[0007] 第三方面,本发明实施例提供一种电子设备,包括:至少一个处理器;以及与所述处理器连接的至少一个存储器、总线;其中,所述处理器、存储器通过所述总线完成相互间的通信;所述处理器用于调用所述存储器中的程序指令,以执行上述一个或多个技术方案中的方法。

[0008] 第四方面,本发明实施例提供一种计算机可读存储介质,所述存储介质包括存储的程序,其中,在所述程序运行时控制所述存储介质所在设备执行上述一个或多个技术方案中的方法。

[0009] 本发明实施例提供的基于长文本的信息推荐方法及装置,首先,对长文本进行意图识别,得到多个短文本,短文本为基于长文本中的分词扩展出的文本;然后,对每一短文本进行意图识别,得到每一短文本的推荐信息;接着,从每一短文本的推荐信息中获取与长文本中的分词关联的推荐信息;最后,向用户推荐与长文本中的分词关联的推荐信息。由于向用户推荐的推荐信息是根据长文本中的分词从每一短文本的推荐信息中获取的,因此,

能够减少推荐信息的数量,以使用户查找,并且,由于每一短文本是通过对长文本进行意图识别得到的,因此,还能够使推荐的信息更加精准。

[0010] 上述说明仅是本发明实施例技术方案的概述,为了能够更清楚了解本发明实施例的技术手段,而可依照说明书的内容予以实施,并且为了让本发明实施例的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明实施例的具体实施方式。

### 附图说明

[0011] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明实施例的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0012] 图1为本发明实施例中的基于长文本的信息推荐方法的流程示意图一;

[0013] 图2为本发明实施例中的基于长文本的信息推荐方法的流程示意图二;

[0014] 图3为本发明实施例中的基于长文本的信息推荐装置的结构示意图;

[0015] 图4为本发明实施例中的电子设备的结构示意图。

### 具体实施方式

[0016] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0017] 本发明实施例提供了一种基于长文本的信息推荐方法,在实际应用中,当用户在电子设备上使用短信、微信等即时通讯类软件进行聊天时,或者在网页或应用程序内浏览新闻时,有时需要基于聊天界面或网页中的某一文本获得推荐信息,通过本发明实施例提供的基于长文本的信息推荐方法,能够基于用户选择的文本向用户推荐该文本的相关信息。尤其是当用户选择的文本为长文本时,能够避免向用户推荐大量的信息,实现信息的精准推荐。

[0018] 下面对本发明实施例提供的基于长文本的信息推荐方法进行详细说明。

[0019] 图1为本发明实施例中的基于长文本的信息推荐方法的流程示意图一,参见图1所示,该方法可以包括:

[0020] S101:对长文本进行意图识别,得到多个短文本。

[0021] 其中,短文本为基于长文本中的分词扩展出的文本。

[0022] 所谓对长文本进行意图识别,就是识别出长文本中的意思。即用户看到例如新闻标题的文本后想要获知的更多信息,对于长文本来说,其中可以是存在有多个意思的,而这每一个意思所表现出来的就是一个短文本。

[0023] 举例来说,对于“短道世锦赛1500米范可新无缘决赛韩国包揽两金”这一长文本,对该长文本进行意图识别,就能够得到多个短文本。例如:“短道世锦赛金牌榜”、“短道世锦赛韩国名将”、“范可新有几块金牌”等文本。当然,得到的上述三个短文本仅为举例,通过对上述长文本进行意图识别,还可以得到其它的短文本,在此不做限定。

[0024] S102:对每一短文本进行意图识别,得到每一短文本的推荐信息。

[0025] 在得到多个短文本后,还需要分别对每一个短文本进行意图识别。即识别每一个短文本的意图。例如:得到的短文本为“国贸大厦”,对该短文本进行意图识别,能够得到该短文本是一个基于地址的意图。再例如:得到的短文本为“明天上午九点”,对该短文本进行意图识别,能够得到该短文本是一个基于时间的意图。再例如:得到的短文本为“金牌榜”,对该短文本进行意图识别,能够得到该短文本是一个基于金牌数量推荐的意图。

[0026] 在识别出每一个短文本的意图后,根据意图的不同,要么是直接基于短文本的意图向用户进行推荐,要么是基于短文本的意图进行搜索,将搜索结果推荐给用户。例如:对于“明天上午九点”这一短文本,识别出该短文本是一个基于时间的意图,该用户可能需要进行记事,因此,就将记事本作为推荐信息。再例如:对于“金牌榜”这一短文本,识别出该短文本是一个基于金牌数量推荐的意图,因此,将“金牌榜”输入搜索引擎中进行搜索,就能够得到各种赛事的金牌获得情况,进而将各种赛事的金牌获得情况作为推荐信息。在这里,搜索引擎可以是百度、谷歌、必应等搜索引擎,当然,还可以是其它搜索引擎,在此不做具体限定。

[0027] S103:从每一短文本的推荐信息中获取与长文本中的分词关联的推荐信息。

[0028] 由于每一个短文本的推荐信息都可能会存在数量较大的问题,这样不利于用户查找,因此,需要再次结合长文本,从每一短文本的推荐信息中筛选出与长文本中的分词存在关联的推荐信息。

[0029] 示例性的,对于“金牌榜”这一短文本,得到的推荐信息的数量就较多,例如:奥运会金牌榜、亚运会金牌榜、世锦赛金牌榜、亚锦赛金牌榜等等,而结合“短道世锦赛1500米范可新无缘决赛韩国包揽两金”这一长文本,就可以从众多的金牌榜中筛选出短道世锦赛金牌榜,进而较少推荐信息的数量,以使用户查找。

[0030] S104:向用户推荐与长文本中的分词关联的推荐信息。

[0031] 在实际应用中,可以在用户当前使用的界面的下半部分界面中插入推荐信息,这样,既不会影响用户对于当前界面的使用,还能够基于用户选择的长文本向用户推荐信息。

[0032] 由上述内容可知,本发明实施例提供的基于长文本的信息推荐方法,首先,对长文本进行意图识别,得到多个短文本,短文本为基于长文本中的分词扩展出的文本;然后,对每一短文本进行意图识别,得到每一短文本的推荐信息;接着,从每一短文本的推荐信息中获取与长文本中的分词关联的推荐信息;最后,向用户推荐与长文本中的分词关联的推荐信息。由于向用户推荐的推荐信息是根据长文本中的分词从每一短文本的推荐信息中获取的,因此,能够减少推荐信息的数量,以使用户查找,并且,由于每一短文本是通过对长文本进行意图识别得到的,因此,还能够使推荐的信息更加精准。

[0033] 进一步地,作为对图1所示方法的细化和扩展,本发明实施例还提供了一种基于长文本的信息推荐方法。图2为本发明实施例中的基于长文本的信息推荐方法的流程示意图二,参见图2所示,该方法可以包括:

[0034] S201:获取长文本中的多个分词。

[0035] 由于长文本中存在有多个分词,并且最终向用户推荐的内容都是基于长文本中是分词得到的,因此,在确定用户选中的长文本后,就需要先获取长文本中的多个分词。

[0036] 在实际应用中,可以通过分词等方法获取长文本中的分词,当然,也可以通过其它方法获取长文本中的分词,在此不做具体限定。

[0037] 在获取到长文本中的多个分词后,就需要基于这多个分词生成多个短文本。在具体实施过程中,可以采用S202和S203中的任意一种方法得到长文本对应的多个短文本。

[0038] S202:基于每一个分词生成一个短文本,将最终生成的所有短文本作为长文本对应的多个短文本。

[0039] 具体来说,就是对长文本中的每一个分词进行意图识别,得到每一个分词对应的短文本,这里每一个分词对应的短文本就形成了长文本对应的多个短文本。

[0040] 举例来说,假设从某长文本中获取到“短道世锦赛”、“两金”这两个分词,接下来,对这两个分词分别进行意图识别,“短道世锦赛”这一分词对应的短文本就是“短道世锦赛”,而“两金”这一分词对应的短文本就是“金牌榜”。短道世锦赛和“金牌榜”这两个短文本就是该长文本对应的多个短文本。

[0041] S203:基于至少两个分词生成一个短文本,将最终生成的所有短文本作为长文本对应的多个短文本。

[0042] 在这里,根据长文本中的分词的数量,可以选择每两个分词生成一个短文本,也可以选择每三个分词生成一个短文本,还可以选择两个分词生成一个短文本和三个分词生成一个短文本,在此不做限定。需要说明的是,上述两个、三个仅为举例,也可以是四个、五个或六个等等,需要根据长文本中的分词的实际数量以及实际情况确定。

[0043] 示例性的,假设某长文本中有五个分词,分别是:分词A、分词B、分词C、分词D和分词E。其中一种生成多个短文本的方式是:基于分词A和分词B生成短文本a,基于分词B和分词C生成短文本b,基于分词A和分词C生成短文本c,基于分词A、分词B和分词C生成短文本d。这样,基于该文本中的至少两个分词,就生成了四个短文本,即短文本a、短文本b、短文本c和短文本d。

[0044] 在这里需要说明的是:在基于长文本中的至少两个分词生成一个短文本的过程中,长文本中的分词可以只使用一次,也可以使用多次,在此不做限定。而为了获得更多的短文本,使短文本的推荐信息更加全面,进而使得推荐的信息更加精准,可以采用排列组合中组合的方式将长文本中的多个分词两两合并,三三合并,直到将多个分词合并在一起生成一个短文本为止。

[0045] 当将长文本中的至少两个分词进行合并,得到多个合并后的分词后,有些合并后的分词并没有对应的推荐信息。例如:将“短道世锦赛”、“1500米”、“范可新”、“金牌榜”这四个分词进行合并,而由于范可新在短道世锦赛1500米中并未夺金,因此将上述四个分词进行合并后的分词并无对应的推荐信息。而为了避免搜索无对应推荐信息的合并后的分词,提高推荐效率,可以通过计算合并后的分词的权重值,在这里,权重值越小,表示分词对应的信息就越少,当某一分词的权重值小于预设阈值时,就说明该分词并无对应的信息。因此,在计算出合并后的分词的权重值后,删除权重值小于预设阈值的合并后的分词,保留权重值大于或等于预设阈值的合并后的分词,并将保留的合并后的分词作为长文本对应的多个短文本,或者对保留的合并后的分词进行扩展,扩展出短文本,将扩展出的短文本作为长文本对应的多个短文本。

[0046] 通过将长文本中的至少两个分词合并,基于合并后的分词生成短文本,能够使短文本的意图更加明确,进而减少短文本的推荐信息的数量,提高基于长文本对短文本的推荐信息的筛选效率,最终提高信息的推荐效率。例如:将“短道世锦赛”和“两金”这两个分词

进行合并,就能够得到“短道世锦赛金牌榜”这一短文本,相比于基于“金牌榜”这一短文本获得推荐信息有奥运会金牌榜、世锦赛金牌榜等,基于“短道世锦赛金牌榜”这一合并后的分词生成短文本获得推荐信息就只有短道世锦赛金牌榜,而并没有奥运会金牌榜和其它世锦赛金牌榜,如此,就减少了短文本的推荐信息的数量。

[0047] 至此,采用S202或S203中的任意一种方法,就能够获得多个短文本了。

[0048] S204:通过分词处理、命名实体识别、语义分析中的一种或多种对每一短文本进行意图识别,得到每一短文本的推荐信息。

[0049] 在上述实施例中已将对每一短文本进行意图识别,得到每一短文本的推荐信息的具体过程进行了说明,故在此不再赘述。接下来主要对如何对每一短文本进行意图识别进行说明。

[0050] 以对一个短文本进行意图识别为例,在具体实施过程中,可以通过以下两种方式中的任意一种方式来对短文本进行意图识别。当然,还可以通过其它方式,在此不做具体限定。

[0051] 第一种方式:先对短文本进行分词处理,得到分词结果,再基于分词结果进行命名实体识别,进而识别出短文本的意图。

[0052] 例如:对于“明天下午3点到国贸大厦302开会”这一短文本,首先,通过分词处理,得到“明天”、“下午”、“3点”、“到”、“国贸大厦”、“302”、“开会”这些分词结果,然后,对这些分词结果进行命名实体识别,得到“明天下午3点”这一基于时间的命名实体、“国贸大厦302”这一基于地点的命名实体、以及“开会”这一基于事件的命名实体,进而识别出需要推荐记事本和地图。这里的命名实体,即是Named Entity Recognition,简称NER。

[0053] 第二种方式:对短文本进行语义分析,识别出短文本的意图。

[0054] 例如:对于“我想去国贸大厦”这一短文本,通过语义分析,能够识别出需要推荐去往国贸大厦的地图。

[0055] 在这里需要说明的是:在识别出短文本的意图后,若能够基于该意图直接向用户进行推荐,则可以不进行搜索。例如:该意图是一个基于时间的意图,则将记事本作为推荐信息,向用户推荐。若不能够基于该意图直接向用户推荐,则需要基于该意图将短文本输入到搜索引擎中进行搜索,将搜索结果作为推荐信息。例如:该意图是一个推荐金牌数量的意图,则基于该意图将“金牌榜”输入到搜索引擎中进行搜索,将推荐金牌数量的信息作为推荐信息,向用户推荐。

[0056] 通过以上两种方式中的任意一种方式,就能够对短文本进行意图识别,得到短文本的推荐信息了。

[0057] S205:确定每一短文本的推荐信息的数量。

[0058] 其中,每一条推荐信息的数量都为1,短文本的推荐信息的数量为多少,就代表该短文本对应有多少条推荐信息。

[0059] S206:若每一短文本的推荐信息的数量大于或等于第一预设数量,则基于长文本中的分词,对每一短文本的推荐信息进行筛选,获得与长文本中的分词关联的推荐信息。

[0060] 在这里,长文本对应多个短文本,若某个短文本的推荐信息的数量大于或等于第一预设数量,则说明该短文本的推荐信息较多,不利于用户查找,因此,就需要基于长文本中的分词,在该短文本的推荐信息中筛选出与长文本中的分词关联的推荐信息,并且确



保筛选出的推荐信息的数量小于第一预设数量。

[0061] 示例性的,对于“金牌榜”这一短文本,该短文本的推荐信息就比较多,有奥运会金牌榜、世锦赛金牌榜、亚运会金牌榜、亚锦赛金牌榜以及国内各种大小型赛事的金牌榜等较多的推荐信息,这些推荐信息的数量很容易就超过了第一预设数量,因此,需要在结合长文本中的“短道世锦赛”这一分词,从上述各种金牌榜中筛选出短道世锦赛金牌榜,这样,就大大减少了“金牌榜”的推荐信息的数量。

[0062] 在此需要说明的是:在基于长文本中的分词,对短文本的推荐信息进行筛选时,为了能够实现有效的筛选,需要确保长文本中的分词与短文本中的分词是不同的两个分词。

[0063] 并且,还需要说明的是:第一预设数量可以根据实际需求进行设置,可以是一个两个,也可以是三个四个,在此不做具体限定。

[0064] 此外,还可能存在一种情况,就是:每一个短文本的信息推荐的数量都小于第一预设数量,但是短文本的数量较多,所以总体上的推荐信息的数量也较多。在这种情况下,还是需要根据长文本中的分词,对多个短文本的推荐信息进行筛选。具体来说,可以先确定多个短文本的数量,若多个短文本的数量大于或等于第二预设数量时,则根据长文本中的分词,对多个短文本的推荐信息进行筛选,进一步避免向用户推荐较多的信息。在这里,第二预设数量与第一预设数量可以是相同的,也可以是不同的,在此不做限定。

[0065] 最后,如果多个短文本的数量也小于第二预设数量,那么就说明多个短文本的推荐信息的数量不多,因此可以无需对多个短文本的推荐信息进行筛选,而是可以直接向用户推荐,能够进一步提高信息推荐效率。

[0066] S207:向用户推荐与长文本中的分词关联的推荐信息。

[0067] 至此,就完成了整个信息推荐的过程。

[0068] 由上述内容可知,本发明实施例提供的基于长文本的信息推荐方法,首先,获取长文本中的多个分词;然后,基于每一个分词生成一个短文本,将最终生成的所有短文本作为长文本对应的多个短文本;或者基于至少两个分词生成一个短文本,将最终生成的所有短文本作为长文本对应的多个短文本;接着,对每一短文本进行意图识别,得到每一短文本的语义片段;再基于每一短文本的语义片段进行搜索,得到每一短文本的搜索结果,并将每一短文本的搜索结果作为每一短文本的推荐信息;再接着,确定每一短文本的推荐信息的数量;若每一短文本的推荐信息的数量大于或等于第一预设数量,则基于长文本中的分词,对每一短文本的推荐信息进行筛选,获得与长文本中的分词关联的推荐信息;最后,向用户推荐与长文本中的分词关联的推荐信息。由于向用户推荐的推荐信息是根据长文本中的分词从每一短文本的推荐信息中获取的,因此,能够减少推荐信息的数量,以使用户查找,并且,每一短文本都是通过对长文本进行意图识别得到的,因此,还能够使推荐的信息更加精准。

[0069] 基于同一发明构思,作为对上述方法的实现,本发明实施例还提供了一种基于长文本的信息推荐装置。图3为本发明实施例中的基于长文本的信息推荐装置的结构示意图,参见图3所示,该装置30可以包括:获取模块301,用于对长文本进行意图识别,得到多个短文本,所述短文本为基于所述长文本中的分词扩展出的文本;识别模块302,用于对每一短文本进行意图识别,得到所述每一短文本的推荐信息;确定模块303,用于从所述每一短文本的推荐信息中获取与所述长文本中的分词关联的推荐信息;推荐模块304,用于向用户推荐与所述长文本中的分词关联的推荐信息。

[0070] 基于前述实施例,所述获取模块,包括:第一获取单元,用于获取所述长文本中的多个分词;第一识别单元,用于对所述长文本中的每一分词进行意图识别,得到所述每一分词对应的短文本,并将所述每一分词对应的短文本作为所述多个短文本。

[0071] 基于前述实施例,所述获取模块,包括:第二获取单元,用于获取所述长文本中的多个分词;合并单元,用于将所述长文本中的至少两个分词进行合并,得到多个合并后的分词;第二识别单元,用于计算所述多个合并后的分词的权重值,并将权重值大于或等于预设阈值的合并后的分词或者根据合并后的分词扩展出的短文本作为所述多个短文本。

[0072] 基于前述实施例,所述识别模块具体用于通过分词处理、命名实体识别、语义分析中的一种或多种对每一短文本进行意图识别,得到所述每一短文本的推荐信息。

[0073] 基于前述实施例,所述确定模块,包括:第一确定单元,用于确定所述每一短文本的推荐信息的数量;第一选择单元,用于若所述每一短文本的推荐信息的数量大于或等于第一预设数量,则基于所述长文本中的分词,对所述每一短文本的推荐信息进行筛选,获得与所述长文本中的分词关联的推荐信息。

[0074] 基于前述实施例,所述确定模块,还包括:第二确定单元,用于若所述每一短文本的推荐信息的数量小于第一预设数量,则确定所述多个短文本的数量;第二选择单元,用于当所述多个短文本的数量大于或等于第二预设数量时,则基于所述长文本中的分词,对所述多个短文本的推荐信息进行筛选,获得与所述长文本中的分词关联的推荐信息。

[0075] 这里需要指出的是:以上装置实施例的描述,与上述方法实施例的描述是类似的,具有同方法实施例相似的有益效果。对于本发明装置实施例中未披露的技术细节,请参照本发明方法实施例的描述而理解。

[0076] 基于同一发明构思,本发明实施例还提供了一种电子设备。图4为本发明实施例中的电子设备的结构示意图,参见图4所示,该电子设备40可以包括:至少一个处理器401;以及与处理器401连接的至少一个存储器402、总线403;其中,处理器401、存储器402通过总线403完成相互间的通信;处理器401用于调用存储器402中的程序指令,以执行上述一个或多个实施例中的方法。

[0077] 这里需要指出的是:以上电子设备实施例的描述,与上述方法实施例的描述是类似的,具有同方法实施例相似的有益效果。对于本发明实施例的电子设备的实施例中未披露的技术细节,请参照本发明方法实施例的描述而理解。

[0078] 基于同一发明构思,本发明实施例还提供了一种计算机可读存储介质,上述计算机可读存储介质包括存储的程序,其中,在程序运行时控制存储介质所在设备执行上述一个或多个实施例中的方法。

[0079] 这里需要指出的是:以上计算机可读存储介质实施例的描述,与上述方法实施例的描述是类似的,具有同方法实施例相似的有益效果。对于本发明实施例的计算机可读存储介质的实施例中未披露的技术细节,请参照本发明方法实施例的描述而理解。

[0080] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0081] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0082] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0083] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0084] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0085] 存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。存储器是计算机可读介质的示例。

[0086] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0087] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0088] 本领域技术人员应明白,本申请的实施例可提供为方法、系统或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0089] 以上仅为本申请的实施例而已,并不用于限制本申请。对于本领域技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、

改进等,均应包含在本申请的权利要求范围之内。

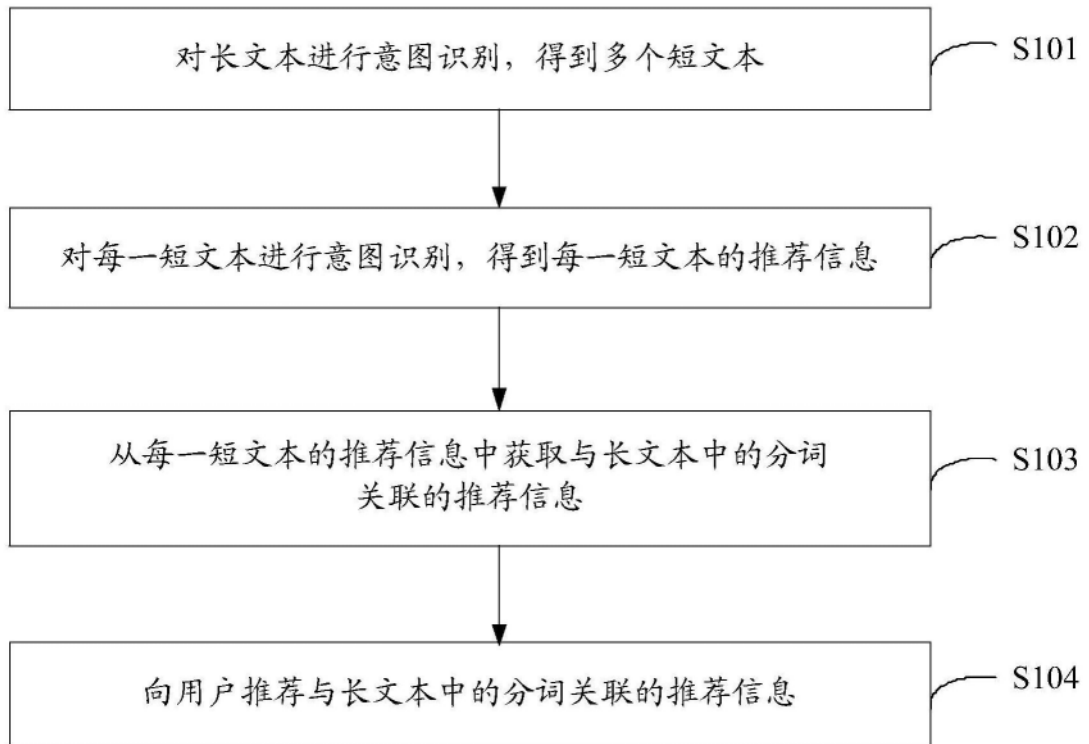


图1

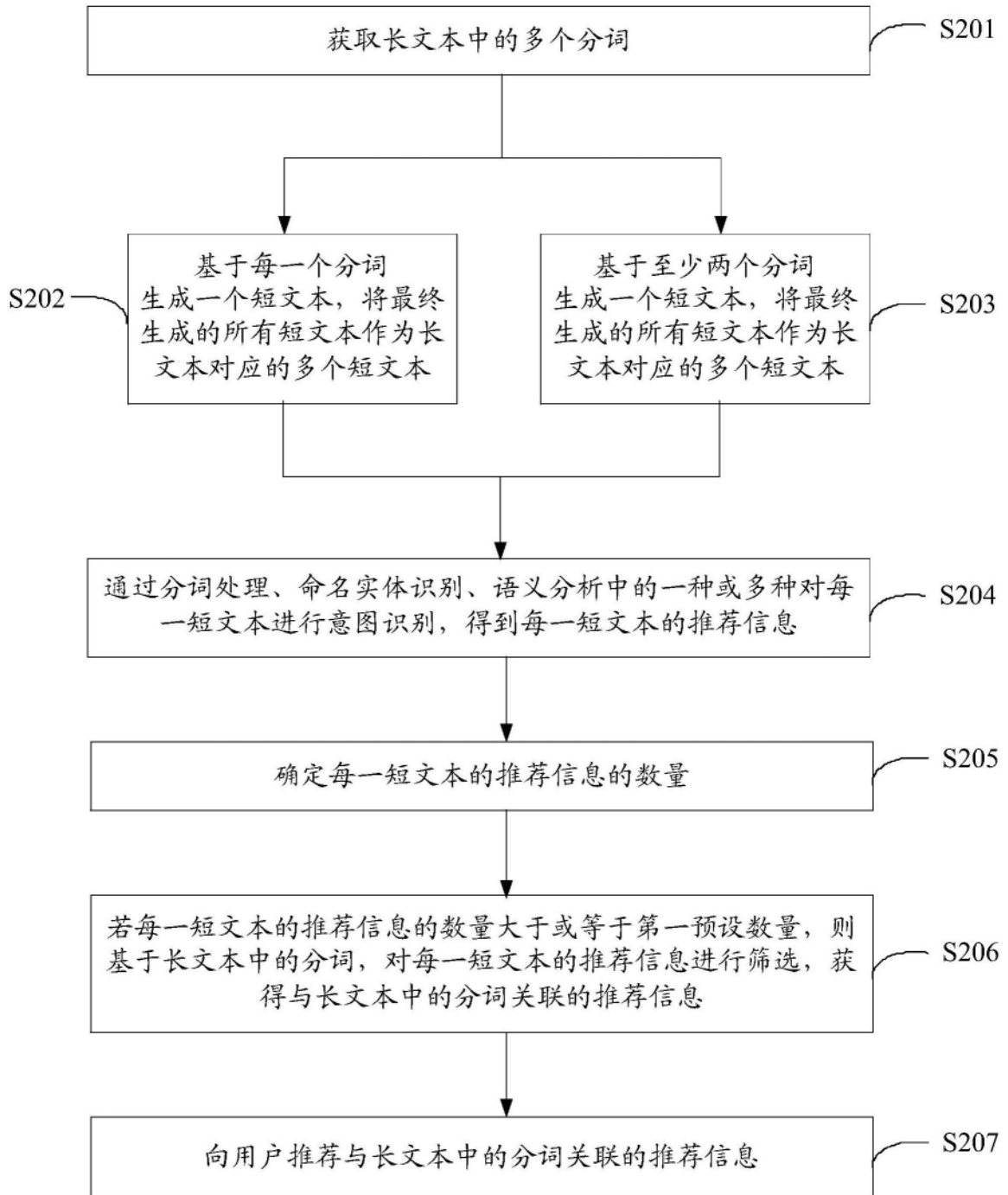


图2

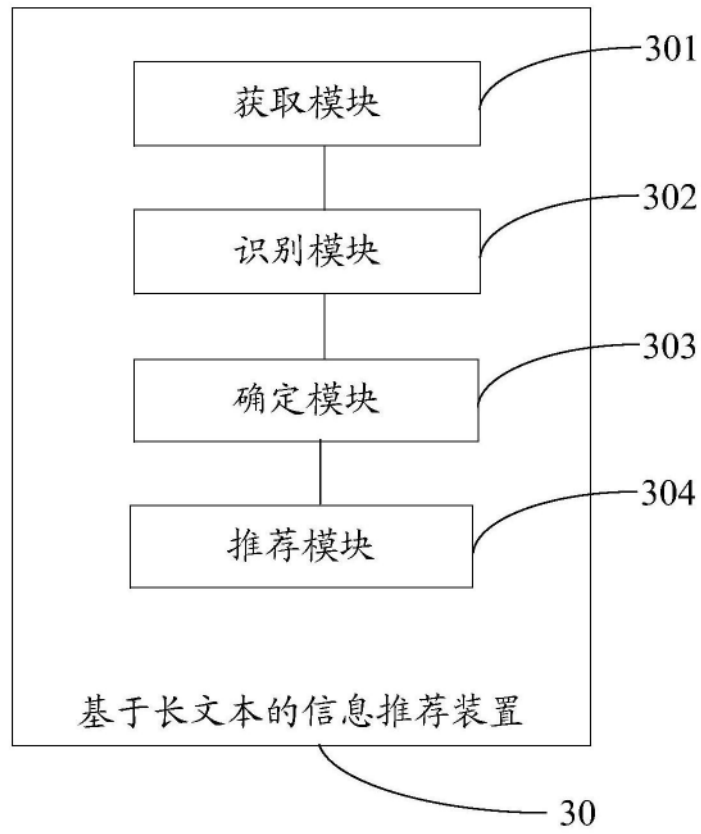


图3

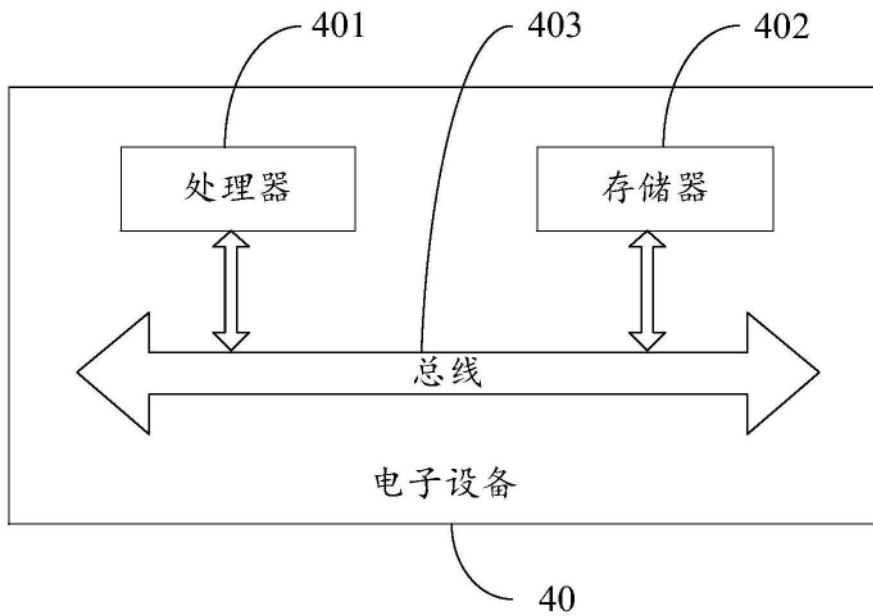


图4