



(12) 发明专利申请

(10) 申请公布号 CN 102750277 A

(43) 申请公布日 2012. 10. 24

(21) 申请号 201110096463. 9

(22) 申请日 2011. 04. 18

(71) 申请人 腾讯科技(深圳)有限公司

地址 518000 广东省深圳市福田区赛格科技园 2 栋东 403 室

(72) 发明人 焦峰 李亚楠 杨月奎

(74) 专利代理机构 北京三高永信知识产权代理有限公司 11138

代理人 刘映东

(51) Int. Cl.

G06F 17/30(2006. 01)

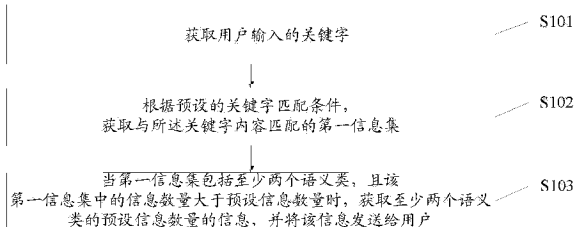
权利要求书 2 页 说明书 9 页 附图 7 页

(54) 发明名称

获取信息的方法和装置

(57) 摘要

本发明公开了一种获取信息的方法,该方法包括:获取用户输入的关键字;根据预设的关键字匹配条件,获取与该关键字内容匹配的第一信息集;判断该第一信息集中的信息数量是否大于预设信息数量,且该第一信息集是否包括至少两个语义类,如果是,则获取该预设信息数量的信息,该信息包括至少两个语义类;发送该信息至该用户。本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。



1. 一种获取信息的方法,其特征在于,所述方法包括:
  - 获取用户输入的关键字;
  - 根据预设的关键字匹配条件,获取与所述关键字内容匹配的第一信息集;
  - 判断所述第一信息集中的信息数量是否大于预设信息数量,且所述第一信息集是否包括至少两个语义类,如果是,则获取所述预设信息数量的信息,所述信息包括至少两个语义类;
  - 发送所述信息至所述用户。
2. 根据权利要求 1 所述的方法,其特征在于,所述判断所述第一信息集中的信息数量是否大于预设信息数量,且所述第一信息集是否包括至少两个语义类,具体包括:
  - 获取所述第一信息集中的信息数量,判断所述信息数量是否大于预设信息数量;
  - 对所述第一信息集中的信息按语义类进行文本聚类;
  - 获取所述第一信息集包含的语义类的数量;
  - 判断所述语义类的数量是否大于或等于两个。
3. 根据权利要求 1 所述的方法,其特征在于,所述获取预设信息数量的信息,所述信息包括至少两个语义类具体包括:
  - 当所述第一信息集包含的语义类的数量小于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第一临时信息集;
  - 计算所述预设信息数量与所述语义类的数量的差值数;
  - 对所述第一信息集中剩余的信息按其与所述关键字的匹配度由高到低进行排序;
  - 获取排序后信息位置序号小于或等于所述差值数的信息,得到第二临时信息集,并将所述第一临时信息集和所述第二临时信息集合并,得到所述预设信息数量的信息;
  - 当所述第一信息集包含的语义类的数量大于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第四临时信息集;
  - 对所述第四临时信息集中的信息按其与所述关键字的匹配度由高到低进行排序;
  - 获取排序后第四临时信息集中信息位置序号小于或等于所述预设信息数量的信息,得到所述预设信息数量的信息。
4. 根据权利要求 1 所述的方法,其特征在于,所述获取预设信息数量的信息,所述信息包括至少两个语义类具体包括:
  - 对所述第一信息集中的信息按其与所述关键字的匹配度由低到高进行排序;
  - 当所述第一信息集为  $SQ = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  为所述第一信息集中的信息数;
  - 则根据  $r_{q_x} = sq_y$  获取至少两个语义类的预设信息数量的信息;
  - 其中,  $y = \lfloor x^a \rfloor$ ,  $a = \log_N m$ ,  $N$  为预设信息数量,  $r_{q_x}$  为按  $y = \lfloor x^a \rfloor$  在  $SQ$  中获取后的信息。
5. 根据权利要求 1 所述的方法,其特征在于,所述将所述信息发送给用户具体包括:
  - 对所述信息按其与所述关键字的匹配度由高到低进行排序;
  - 将排序后的信息,按顺序依次发送给用户。
6. 一种获取信息的装置,其特征在于,所述装置包括:
  - 关键字获取模块,用于获取用户输入的关键字;
  - 第一信息集获取模块,用于根据预设的关键字匹配条件,获取与所述关键字内容匹配

的第一信息集；

信息获取模块,用于判断所述第一信息集中的信息数量是否大于预设信息数量,且所述第一信息集是否包括至少两个语义类,如果是,则获取所述预设信息数量的信息,所述信息包括至少两个语义类；

信息发送模块,用于发送所述信息至所述用户。

7. 根据权利要求6所述的装置,其特征在于,所述信息获取模块,具体包括：

信息数量确定单元,用于获取所述第一信息集中的信息数量,判断所述信息数量是否大于预设信息数量；

文本聚类单元,用于对所述第一信息集中的信息按语义类进行文本聚类；

语义类数量获取单元,用于获取所述第一信息集包含的语义类的数量；

语义类确定单元,用于判断所述语义类的数量是否大于或等于两个；

信息获取单元,用于当所述第一信息集中的信息数量大于预设信息数量,且所述第一信息集包括至少两个语义类时,获取所述预设信息数量的信息,所述信息包括至少两个语义类。

8. 根据权利要求6所述的装置,其特征在于,所述信息获取模块,具体包括：

临时信息集生成单元,用于当所述第一信息集包含的语义类的数量小于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第一临时信息集；

数量差值数计算单元,用于计算所述预设信息数量与所述语义类的数量的差值数；

预设信息获取单元,用于对所述第一信息集中剩余的信息按其与所述关键字的匹配度由高到低进行排序,获取排序后信息位置序号小于或等于所述差值数的信息,得到第二临时信息集,并将所述第一临时信息集和所述第二临时信息集合并,得到所述预设信息数量的信息；

第一信息获取单元,用于当所述第一信息集包含的语义类的数量大于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第四临时信息集,对所述第四临时信息集中的信息按其与所述关键字的匹配度由高到低进行排序,获取排序后第四临时信息集中信息位置序号小于或等于所述预设信息数量的信息,得到所述预设信息数量的信息。

9. 根据权利要求6所述的装置,其特征在于,所述信息获取模块具体包括：

第二信息获取单元,用于对所述第一信息集中的信息按其与所述关键字的匹配度由低到高进行排序,当所述第一信息集为  $SQ = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  为所述第一信息集中的信息数时,则根据  $r_{q_x} = sq_y$  获取至少两个语义类的预设信息数量的信息；其中,  $y = \lfloor x^a \rfloor$ ,  $a = \log_N m$ ,  $N$  为预设信息数量,  $r_{q_x}$  为按  $y = \lfloor x^a \rfloor$  在  $SQ$  中获取后的信息。

10. 根据权利要求6所述的装置,其特征在于,所述信息发送模块,具体包括：

关键字排序单元,用于对所述信息按其与所述关键字的匹配度由高到低进行排序；

信息发送单元,用于将排序后的信息,按顺序依次发送给用户。

## 获取信息的方法和装置

### 技术领域

[0001] 本发明涉及通信技术领域,特别涉及一种获取信息的方法和装置。

### 背景技术

[0002] 问答系统是互联网用户获取信息的一种常见工具,例如百度知道、搜搜问问等。为了满足用户的信息浏览需求,问答系统会检索推送与当前浏览问题相关的其他问题或答案,这里称之为“相关问题”。相关问题可以进一步的满足用户的浏览需求。然而由于受展示空间限制,一个问题的相关问题往往只能显示 5 条左右,很多时候无法将所有的相关问题进行展示,因此需要一定的方法选取最具代表性的几条相关问题。

[0003] 现有相关问题检索系统选取与当前浏览问题语义最接近的几条问题,并依次展示给用户。其技术实现为:首先,获取用户点击或输入的问题 Q;然后,利用信息检索或自然语言处理技术,从以往收集或记录的问题数据库中检索出与问题 Q 相关的问题集 R(Q);然后,对 R(Q) 中相关问题按照与 Q 的语义相关度进行排序;最后,选取 R(Q) 中排名最高的 N 条相关问题进行展示。其中,N 为相关问题在页面上显示条数的最大值。

[0004] 现有技术为用户提供与用户提出的问题内容上相关信息。但通过现有的技术方案给出的相关问题搜索结果,都是语义相同或非常接近的,当用户希望泛泛地浏览某类问题的其他方面的相关知识时,现有的技术方案,无法满足用户需求,需要用户重新输入该某类问题的其他方面信息,重新搜索,减少了用户体验。

[0005] 例如,某用户想装修自己的房子,则该用户可通过输入问题浏览问题为“近几年最流行装修风格是什么?”获取相关内容,相关问题一般只能显示 5 条左右,则通过现有技术方案,用户可获取 5 条关于“装修风格”的信息;但用户可能还希望获取有关装修材料、装修价格、附近装修商口碑等各类知识的相关问题及回答,则需要重新输入需要获取信息的关键字,增加了用户的操作。

### 发明内容

[0006] 为了简化搜索操作,提高用户体验,本发明实施例提供了一种获取信息的方法,所述方法包括:

[0007] 获取用户输入的关键字;

[0008] 根据预设的关键字匹配条件,获取与所述关键字内容匹配的第一信息集;

[0009] 判断所述第一信息集中的信息数量是否大于预设信息数量,且所述第一信息集是否包括至少两个语义类,如果是,则获取所述预设信息数量的信息,所述信息包括至少两个语义类;

[0010] 发送所述信息至所述用户。

[0011] 所述判断所述第一信息集中的信息数量是否大于预设信息数量,且所述第一信息集是否包括至少两个语义类,具体包括:

[0012] 获取所述第一信息集中的信息数量,判断所述信息数量是否大于预设信息数量;

- [0013] 对所述第一信息集中的信息按语义类进行文本聚类；
- [0014] 获取所述第一信息集包含的语义类的数量；
- [0015] 判断所述语义类的数量是否大于或等于两个。
- [0016] 所述获取预设信息数量的信息，所述信息包括至少两个语义类具体包括：
- [0017] 当所述第一信息集包含的语义类的数量小于所述预设信息数量时，则在每个语义类包含的信息中获取一个信息，得到第一临时信息集；
- [0018] 计算所述预设信息数量与所述语义类的数量的差值数；
- [0019] 对所述第一信息集中剩余的信息按其与所述关键字的匹配度由高到低进行排序；
- [0020] 获取排序后信息位置序号小于或等于所述差值数的信息，得到第二临时信息集，并将所述第一临时信息集和所述第二临时信息集合并，得到所述预设信息数量的信息；
- [0021] 当所述第一信息集包含的语义类的数量大于所述预设信息数量时，则在每个语义类包含的信息中获取一个信息，得到第四临时信息集；
- [0022] 对所述第四临时信息集中的信息按其与所述关键字的匹配度由高到低进行排序；
- [0023] 获取排序后第四临时信息集中信息位置序号小于或等于所述预设信息数量的信息，得到所述预设信息数量的信息。
- [0024] 所述获取预设信息数量的信息，所述信息包括至少两个语义类具体包括：
- [0025] 对所述第一信息集中的信息按其与所述关键字的匹配度由低到高进行排序；
- [0026] 当所述第一信息集为  $SQ = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ， $m$  为所述第一信息集中的信息数；
- [0027] 则根据  $r_{q_x} = sq_y$  获取至少两个语义类的预设信息数量的信息；
- [0028] 其中， $y = \lfloor x^a \rfloor$ ， $a = \log_N m$ ， $N$  为预设信息数量， $r_{q_x}$  为按  $y = \lfloor x^a \rfloor$  在  $SQ$  中获取后的信息。
- [0029] 所述将所述信息发送给用户具体包括：
- [0030] 对所述信息按其与所述关键字的匹配度由高到低进行排序；
- [0031] 将排序后的信息，按顺序依次发送给用户。
- [0032] 本发明实施例提供了一种获取信息的装置，所述装置包括：
- [0033] 关键字获取模块，用于获取用户输入的关键字；
- [0034] 第一信息集获取模块，用于根据预设的关键字匹配条件，获取与所述关键字内容匹配的第一信息集；
- [0035] 信息获取模块，用于判断所述第一信息集中的信息数量是否大于预设信息数量，且所述第一信息集是否包括至少两个语义类，如果是，则获取所述预设信息数量的信息，所述信息包括至少两个语义类；
- [0036] 信息发送模块，用于发送所述信息至所述用户。
- [0037] 所述信息获取模块，具体包括：
- [0038] 信息数量确定单元，用于获取所述第一信息集中的信息数量，判断所述信息数量是否大于预设信息数量；
- [0039] 文本聚类单元，用于对所述第一信息集中的信息按语义类进行文本聚类；

- [0040] 语义类数量获取单元,用于获取所述第一信息集包含的语义类的数量;
- [0041] 语义类确定单元,用于判断所述语义类的数量是否大于或等于两个;
- [0042] 信息获取单元,用于当所述第一信息集中的信息数量大于预设信息数量,且所述第一信息集包括至少两个语义类时,获取所述预设信息数量的信息,所述信息包括至少两个语义类。
- [0043] 所述信息获取模块,具体包括:
- [0044] 临时信息集生成单元,用于当所述第一信息集包含的语义类的数量小于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第一临时信息集;
- [0045] 数量差值数计算单元,用于计算所述预设信息数量与所述语义类的数量的差值数;
- [0046] 预设信息获取单元,用于对所述第一信息集中剩余的信息按其与所述关键字的匹配度由高到低进行排序,获取排序后信息位置序号小于或等于所述差值数的信息,得到第二临时信息集,并将所述第一临时信息集和所述第二临时信息集合并,得到所述预设信息数量的信息;
- [0047] 第一信息获取单元,用于当所述第一信息集包含的语义类的数量大于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第四临时信息集,对所述第四临时信息集中的信息按其与所述关键字的匹配度由高到低进行排序,获取排序后第四临时信息集中信息位置序号小于或等于所述预设信息数量的信息,得到所述预设信息数量的信息。
- [0048] 所述信息获取模块具体包括:
- [0049] 第二信息获取单元,用于对所述第一信息集中的信息按其与所述关键字的匹配度由低到高进行排序,当所述第一信息集为  $SQ = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  为所述第一信息集中的信息数时,则根据  $r_{q_x} = sq_y$  获取至少两个语义类的预设信息数量的信息;其中,  $y = \lfloor x^a \rfloor$ ,  $a = \log_N m$ ,  $N$  为预设信息数量,  $r_{q_x}$  为按  $y = \lfloor x^a \rfloor$  在  $SQ$  中获取后的信息。
- [0050] 所述信息发送模块,具体包括:
- [0051] 关键字排序单元,用于对所述信息按其与所述关键字的匹配度由高到低进行排序;
- [0052] 信息发送单元,用于将排序后的信息,按顺序依次发送给用户。
- [0053] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,为用户提供与关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

#### 附图说明

- [0054] 图 1 是本发明实施例 1 提供的获取信息的方法流程图;
- [0055] 图 2 是本发明实施例 2 提供的获取信息的方法流程图;
- [0056] 图 3 是本发明实施例 3 提供的获取信息的方法流程图;
- [0057] 图 4 是本发明实施例 4 提供的获取信息的方法流程图;
- [0058] 图 5 是本发明实施例 5 提供的获取信息的装置示意图;

- [0059] 图 6 是本发明实施例 6 提供的获取信息的装置示意图；  
[0060] 图 7 是本发明实施例 7 提供的获取信息的装置示意图；  
[0061] 图 8 是本发明实施例 8 提供的获取信息的装置示意图。

## 具体实施方式

[0062] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

### [0063] 实施例 1

[0064] 如图 1 所示，本发明实施例提供了一种获取信息的方法，所述方法包括：

[0065] S101：获取用户输入的关键字；

[0066] S102：根据预设的关键字匹配条件，获取与所述关键字内容匹配的第一信息集；

[0067] S103：当所述第一信息集包括至少两个语义类，且所述第一信息集中的信息数量大于预设信息数量时，获取至少两个语义类的预设信息数量的信息，并将所述信息发送给用户。

[0068] 需要说明的是，本发明实施例各步骤的执行主体可以为搜索服务器，也可以为具有个步骤功能的其他执行主体。

[0069] 本发明实施例，在与用户输入的关键字匹配的信息中，获取至少两个语义类的信息，从而为用户提供与其提供的关键字相关类型的信息，从而无需用户重新输入与该关键字相关的关键字，即可获取相关信息，减少了用户的操作，提高了用户体验。

### [0070] 实施例 2

[0071] 如图 2 所示，本发明实施例提供了一种获取信息的方法，所述方法包括：

[0072] S201：获取用户输入的关键字；

[0073] 其中，用户输入的关键字可以是用户提问输入的问题、用户搜索输入的查询、或用户将要浏览的能反映其信息需求的已有问题。

[0074] 例如，通过获取用户输入的问题，获取到用户输入的问题  $q_i$ 。

[0075] S202：根据预设的关键字匹配条件，获取与所述关键字内容匹配的第一信息集；

[0076] 可选的，可通过现有技术进行信息检索技术，从现有的问答系统以往收集和 / 或记录的问题信息数据库中检索出所有与用户问题  $q_i$  语义相关的问题。

[0077] 例如，通过在数据库中对问题  $q_i$  进行检索，得到相关问题候选集  $SQ_i = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ 。

[0078] S203：获取所述第一信息集中的信息数量，判断所述信息数量是否大于预设信息数量，如果是，则执行 S204，如果不是，则将该问题获选集中的信息作为返回给用户的信息，即执行 S206；

[0079] 可选的，当 S202 中相关问题候选集  $SQ_i = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ， $m$  取值为 20，且预设信息数量为 10，即第一信息集中的信息数量大于预设信息数量，则执行 S204。

[0080] S204：对所述第一信息集进行文本聚类；

[0081] 其中，文本聚类主要是依据聚类假设，即同类的文档相似度较大，而不同类的文档相似度较小。

[0082] 优选的，对搜索引擎返回的结果进行聚类，使用户迅速定位到所需要的信息。具体

的,通过用户输入检索关键词,而后对检索到的文档进行聚类处理,并输出各个不同类别的简要描述,从而可以缩小检索的范围,用户只需关注比较有希望的主题。另外这种方法也可以为用户二次检索提供线索。

[0083] 可选的,对所述第一信息集进行文本聚类的算法课包括:划分法(partitioning methods)、层次法(hierarchical methods)、基于密度的方法(density-based methods)、基于网格的方法(grid-based methods)和基于模型的方法(model-based methods)。

[0084] 其中,划分法(partitioning methods)是指:给定一个有N个元组或者纪录的数据集,分裂法将构造K个分组,每一个分组就代表一个聚类, $K < N$ 。而且这K个分组满足下列条件:(1)每一个分组至少包含一个数据纪录;(2)每一个数据纪录属于且仅属于一个分组(注意:这个要求在某些模糊聚类算法中可以放宽);对于给定的K,算法首先给出一个初始的分组方法,以后通过反复迭代的方法改变分组,使得每一次改进之后的分组方案都较前一次好,而所谓好的标准就是:同一分组中的记录越近越好,而不同分组中的纪录越远越好。使用这个基本思想的算法有:K-MEANS 算法、K-MEDOIDS 算法、CLARANS 算法。

[0085] 层次法(hierarchical methods)是指:对给定的数据集进行层次似的分解,直到某种条件满足为止。具体又可分为“自底向上”和“自顶向下”两种方案。例如在“自底向上”方案中,初始时每一个数据纪录都组成一个单独的组,在接下来的迭代中,它把那些相互邻近的组合成一个组,直到所有的记录组成一个分组或者某个条件满足为止。代表算法有:BIRCH 算法、CURE 算法、CHAMELEON 算法等。

[0086] 基于密度的方法(density-based methods)是指:基于密度的方法与其它方法的一个根本区别是:它不是基于各种各样的距离的,而是基于密度的。这样就能克服基于距离的算法只能发现“类圆形”的聚类的缺点。这个方法的指导思想就是,只要一个区域中的点的密度大过某个阈值,就把它加到与之相近的聚类中去。代表算法有:DBSCAN 算法、OPTICS 算法、DENCLUE 算法等。

[0087] 基于网格的方法(grid-based methods)是指:首先将数据空间划分成为有限个单元(cell)的网格结构,所有的处理都是以单个的单元为对象的。这么处理的一个突出的优点就是处理速度很快,通常这是与目标数据库中记录的个数无关的,它只与把数据空间分为多少个单元有关。代表算法有:STING 算法、CLIQUE 算法、WAVE-CLUSTER 算法。

[0088] 基于模型的方法(model-based methods)是指:基于模型的方法给每一个聚类假定一个模型,然后去寻找能个很好的满足这个模型的数据集。这样一个模型可能是数据点在空间中的密度分布函数或者其它。它的一个潜在的假定就是:目标数据集是由一系列的概率分布所决定的。通常有两种尝试方向:统计的方案和神经网络的方案。

[0089] 本步骤中还可通过其他算法实现对第一信息集中的数据进行聚类,本实施例并不限定。

[0090] S205:获取所述第一信息集包含的语义类的数量;

[0091] 例如,对相关问题候选集  $SQ_1 = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ , m 取值为 20 的第一信息候选集按其语义类进行聚类,得到 3 个语义类。

[0092] S206:判断所述语义类的数量是否大于或等于两个,如果大于,则获取预设信息数量的信息,所述信息至少为两个语义类。

[0093] 例如,如 S205 中示例所示,该第一信息集的语言类为 3 类,大于两个语义类,则获



取预设信息数量的信息,所述信息至少为两个语义类。

[0094] S207:对所述信息按其与所述关键字的匹配度由高到低进行排序;

[0095] S208:将排序后的信息,按顺序一次发送给用户。

[0096] 需要说明的是,本发明实施例各步骤的执行主体可以为搜索服务器,也可以为具有个步骤功能的其他执行主体。

[0097] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0098] 实施例 3

[0099] 如图 3 所示,本发明实施例提供了一种获取信息的方法,所述方法包括步骤 S301 ~ S310,其中 S301 ~ S305 与实施例 2 中的 S201 ~ S205 相同,此处不再赘述,与实施例 2 不同的是,本实施里还包括以下步骤:

[0100] S306:判断所述第一信息集包含的语义类的数量是否大于所述预设信息数量,如果大于,则在每个语义类包含的信息中获取一个信息,得到第一临时信息集;

[0101] 例如,相关问题候选集  $SQ_i = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  取值为 20 中包含的语义类的数量为 3,预设信息数量为 10;则在每一个语义类中获取一个信息,此处得到 3 个不同语义类的信息,组成第一临时信息集  $LQ1 = \{lq1_0, lq1_1, lq1_2\}$ 。

[0102] S307:计算所述预设信息数量与所述语义类的数量的差值数;

[0103] 例如,在 S306 获取到 3 个信息后,则计算所述预设信息数量与所述语义类的数量的差值数,即预设信息数量 10 减去 3,差值数为 7。

[0104] S308:对所述第一信息集中剩余的信息按其与所述关键字的匹配度由高到低进行排序;

[0105] 例如,相关问题候选集  $SQ_i = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  取值为 20 中,出去已获取的 3 个信息,还有 17 个信息,对该 17 个信息按其与关键字的匹配度由高到低进行排序。

[0106] S309:获取排序后信息位置序号小于或等于所述差值数的信息,得到第二临时信息集,并将所述第一临时信息集和所述第二临时信息集合并;

[0107] 例如,排序后的剩余信息序号为 1 ~ 17,则获取位置序号等于或小于差值数 7 的信息,即信息序号为 1 ~ 7 的信息,从而得到第二临时信息集  $LQ2 = \{lq2_0, lq2_1, lq2_2, lq2_3, lq2_4, lq2_5\}$ ,并将第一临时信息集与第二临时信息集合并。

[0108] S310:将合并后的信息发送给用户。

[0109] 例如,将第一临时信息集与第二临时信息集合并得到信息  $lq1_0, lq1_1, lq1_2, lq2_0, lq2_1, lq2_2, lq2_3, lq2_4, lq2_5$ ,将该信息发送给用户。

[0110] 优选的,还可对该信息按其与关键字的匹配度由高到低进行排序,将排序后的信息,按顺序发送给用户。

[0111] 需要说明的是本实施例仅为一种获取不同语义类的信息获取方法,具体的获取到不同语义类的信息还可通过多种方法实现,以实现获取到的信息属于不同语义类为目的采用的方法都属于本实施例保护的范围,具体不再赘述。

[0112] 需要说明的是,本发明实施例各步骤的执行主体可以为搜索服务器,也可以为具有个步骤功能的其他执行主体。

[0113] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0114] 实施例 4

[0115] 如图 4 所示,本发明实施例提供了一种获取信息的方法,所述方法包括:

[0116] S401:获取用户输入的关键字;

[0117] 其中,用户输入的关键字可以是用户提问输入的问题、用户搜索输入的查询、或用户将要浏览的能反映其信息需求的已有问题。

[0118] 例如,通过获取用户输入的问题,获取到用户输入的问题  $q_i$ 。

[0119] S402:根据预设的关键字匹配条件,获取与所述关键字内容匹配的第一信息集;

[0120] 可选的,可通过现有技术进行信息检索技术,从现有的问答系统以往收集和/或记录的问题信息数据库中检索出所有与用户问题  $q_i$  语义相关的问题。

[0121] S403:获取所述第一信息集中的信息数量,判断所述信息数量是否大于预设信息数量,如果大于,则执行 S404,如果小于,则执行 S405;

[0122] 本实施例中,可选的,当所述信息数量大于预设信息数量时,可对所述第一信息集中的信息按其与所述关键字的匹配度由低到高进行排序后执行 S404。

[0123] 例如,通过现有技术进行信息检索技术,从现有的问答系统以往收集和/或记录的问题信息数据库中检索出所有与用户问题  $q_i$  语义相关的问题后,根据它们与问题  $q_i$  的相似度进行排序得到相关问题候选集  $SQ_i = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ 。

[0124] S404:当所述第一信息集为  $SQ = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  为所述第一信息集中的信息数;则根据  $r_{q_x} = sq_y$  获取至少两个语义类的预设信息数量的信息。

[0125] 其中,  $y = \lfloor x^a \rfloor$ ,  $a = \log_N m$ ,  $N$  为预设信息数量,  $r_{q_x}$  为按  $y = \lfloor x^a \rfloor$  在  $SQ$  中获取后的信息。

[0126] 具体的,从  $SQ_i$  中取出  $N$  个语义逐步发散的相关问题。令  $N^a = m$ , 即  $a = \log_N m$ , 取函数  $y = \lfloor x^a \rfloor$ , 则  $sq_y$  即为第  $x$  个相关问题  $r_{q_x}$ , 从而得到获取后的信息集  $RQ_i = \{rq_1, rq_2, \dots\}$ 。 $x$  到  $y$  是一个逐步发散的非线性映射,这样既能保证优先输出序列  $SQ_i$  中与  $q_i$  最相关的查询,也能保证  $SQ_i$  后面的语义相关但发散问题能输出到相关问题中。

[0127] 可选的,还可对  $SQ_i$  进行排序后,构造映射函数  $y = f(x)$ , ( $f(N) \leq m$ ), 令  $r_{q_x} = sq_y$ , 从而获得关键问题  $RQ_i = \{rq_1, rq_2, \dots, rq_N\}$ 。各种合适的映射函数  $f(x)$  均可以用于解决该问题,如幂函数、指数函数等。

[0128] S405:将所述信息发送给用户。

[0129] 可选的,输出问题  $q_i$  的相关问题  $RQ_i = \{rq_1, rq_2, \dots\}$ , 在问题浏览页面向用户依次展示各个相关问题。

[0130] 需要说明的是,本发明实施例各步骤的执行主体可以为搜索服务器,也可以为具有个步骤功能的其他执行主体。

[0131] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0132] 实施例 5

[0133] 如图 5 所示,本发明实施例提供了一种获取信息的装置,所述装置包括:关键字获取模块 501、第一信息集获取模块 502、信息获取模块 503 和信息发送模块 504,其中:

[0134] 关键字获取模块 501,用于获取用户输入的关键字;

[0135] 第一信息集获取模块 502,用于根据预设的关键字匹配条件,获取与所述关键字内容匹配的第一信息集;

[0136] 信息获取模块 503,用于判断所述第一信息集中的信息数量是否大于预设信息数量,且所述第一信息集是否包括至少两个语义类,如果是,则获取所述预设信息数量的信息,所述信息包括至少两个语义类;

[0137] 信息发送模块 504,用于发送所述信息至所述用户。

[0138] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0139] 实施例 6

[0140] 如图 6 所示,本发明实施例提供了一种获取信息的装置,与实施例 5 相似,所述装置包括关键字获取模块 501、第一信息集获取模块 502、信息获取模块 503 和信息发送模块 504。

[0141] 进一步所述信息获取模块 503 具体包括:

[0142] 信息数量确定单元 5031,用于获取所述第一信息集中的信息数量,判断所述信息数量是否大于预设信息数量,如果大于,则所述第一信息集中的信息数量大于预设信息数量;

[0143] 文本聚类单元 5032,用于对所述第一信息集中的信息按语义类进行文本聚类;

[0144] 语义类数量获取单元 5033,用于获取所述第一信息集包含的语义类的数量;

[0145] 语义类确定单元 5034,用于判断所述语义类的数量是否大于或等于两个,如果大于,则所述第一信息集包括至少两个语义类。

[0146] 信息获取单元 5035,用于当所述第一信息集中的信息数量大于预设信息数量,且所述第一信息集包括至少两个语义类时,获取所述预设信息数量的信息,所述信息包括至少两个语义类。

[0147] 其中,所述信息发送模块 504,具体包括:

[0148] 关键字排序单元 5041,用于对所述信息按其与所述关键字的匹配度由高到低进行排序;

[0149] 信息发送单元 5042,用于将排序后的信息,按顺序依次发送给用户。

[0150] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0151] 实施例 7

[0152] 如图 7 所示,本发明实施例提供了一种获取信息的装置,与实施例 6 相似,所述装置包括:关键字获取模块 501、第一信息集获取模块 502、信息获取模块 503 和信息发送模块 504。其中,信息发送模块 504 包括:关键字排序单元 5041 和信息发送单元 5042,与实施例

6 不同的是,本实施例中,信息获取模块 503 具体包括:

[0153] 临时信息集生成单元 5036,用于当所述第一信息集包含的语义类的数量小于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第一临时信息集;

[0154] 数量差值数计算单元 5037,用于计算所述预设信息数量与所述语义类的数量的差值数;

[0155] 预设信息获取单元 5038,用于对所述第一信息集中剩余的信息按其与所述关键字的匹配度由高到低进行排序,获取排序后信息位置序号小于或等于所述差值数的信息,得到第二临时信息集,并将所述第一临时信息集和所述第二临时信息集合并,得到所述预设信息数量的信息;

[0156] 第一信息获取单元 5039,用于当所述第一信息集包含的语义类的数量大于所述预设信息数量时,则在每个语义类包含的信息中获取一个信息,得到第四临时信息集,对所述第四临时信息集中的信息按其与所述关键字的匹配度由高到低进行排序,获取排序后第四临时信息集中信息位置序号小于或等于所述预设信息数量的信息,得到所述预设信息数量的信息。

[0157] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0158] 实施例 8

[0159] 如图 8 所示,本发明实施例提供了一种获取信息的装置,与实施例 6 相似,所述装置包括:关键字获取模块 501、第一信息集获取模块 502、信息获取模块 503 和信息发送模块 504。其中,信息发送模块 504 包括:关键字排序单元 5041 和信息发送单元 5042,与实施例 6 不同的是,本实施例中,信息获取模块 503 具体包括:

[0160] 第二信息获取单元 50310,用于对所述第一信息集中的信息按其与所述关键字的匹配度由低到高进行排序,当所述第一信息集为  $SQ = \{sq_0, sq_1, sq_2, \dots, sq_m\}$ ,  $m$  为所述第一信息集中的信息数时,则根据  $r_{q_x} = sq_y$  获取至少两个语义类的预设信息数量的信息;其中,  $y = \lfloor x^a \rfloor$ ,  $a = \log_N m$ ,  $N$  为预设信息数量,  $r_{q_x}$  为按  $y = \lfloor x^a \rfloor$  在  $SQ$  中获取后的信息。

[0161] 本发明实施例,在与用户输入的关键字匹配的信息中,获取至少两个语义类的信息,从而为用户提供与其提供的关键字相关类型的信息,从而无需用户重新输入与该关键字相关的关键字,即可获取相关信息,减少了用户的操作,提高了用户体验。

[0162] 以上实施例提供的技术方案中的全部或部分内容可以通过软件编程实现,其软件程序存储在可读取的存储介质中,存储介质例如:计算机中的硬盘、光盘或软盘。

[0163] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

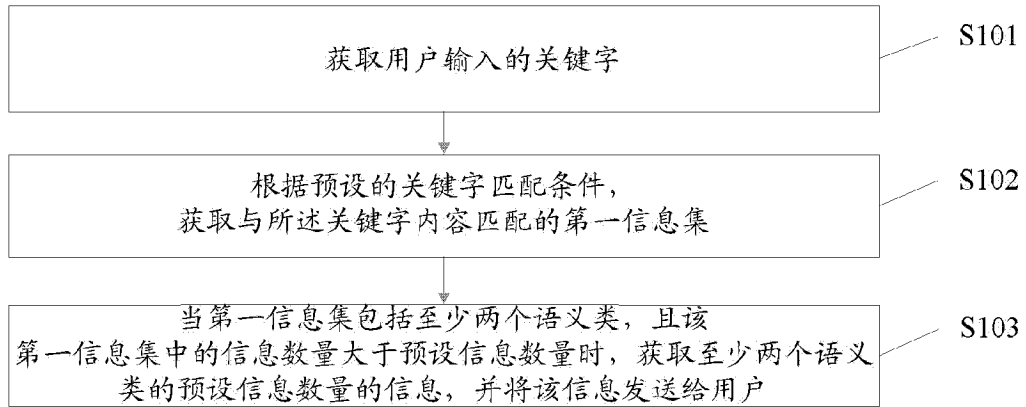


图 1



图2

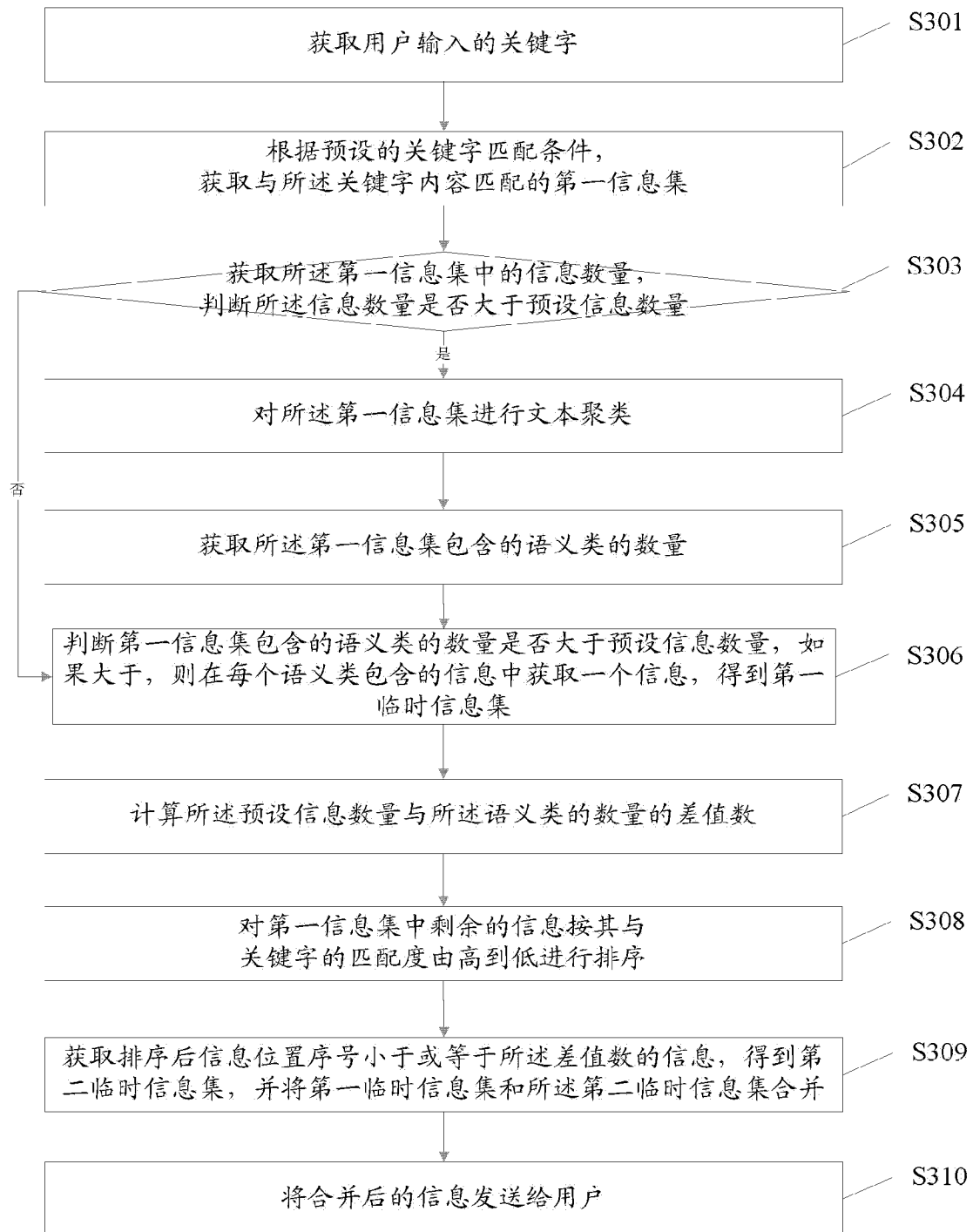


图 3

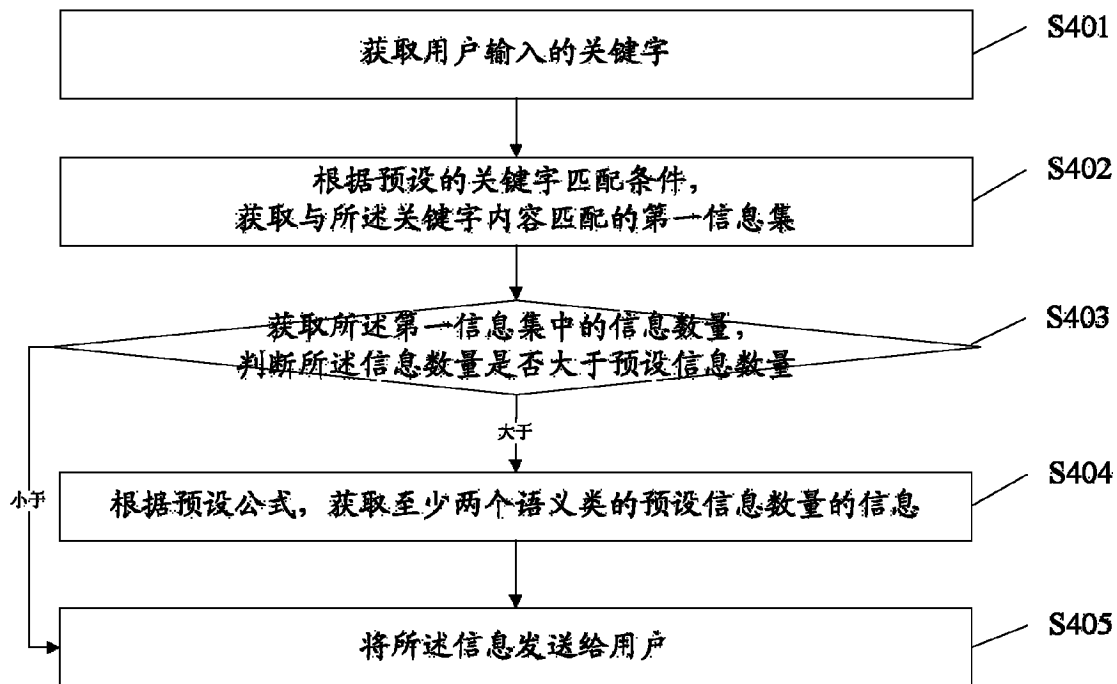


图 4

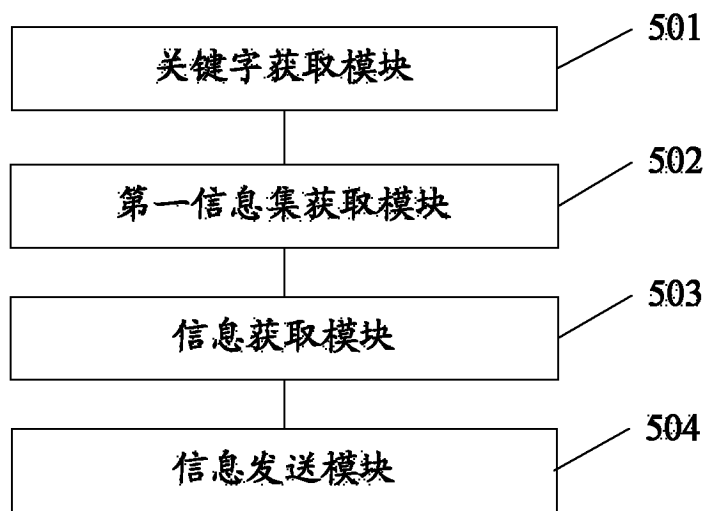


图 5



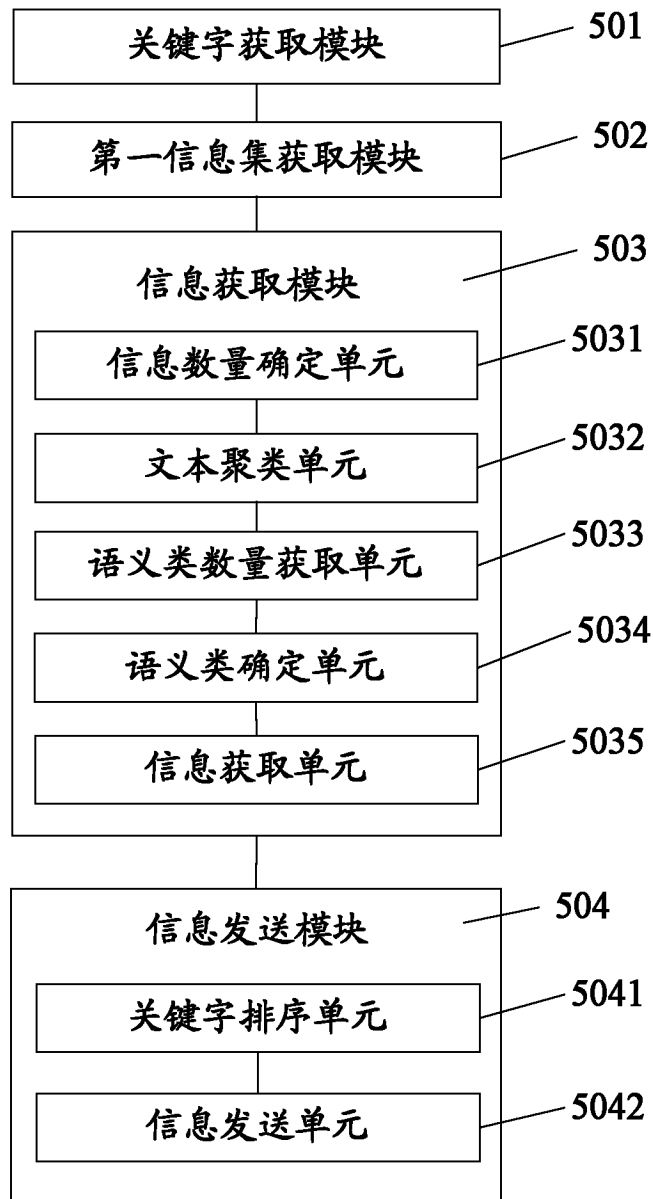


图 6

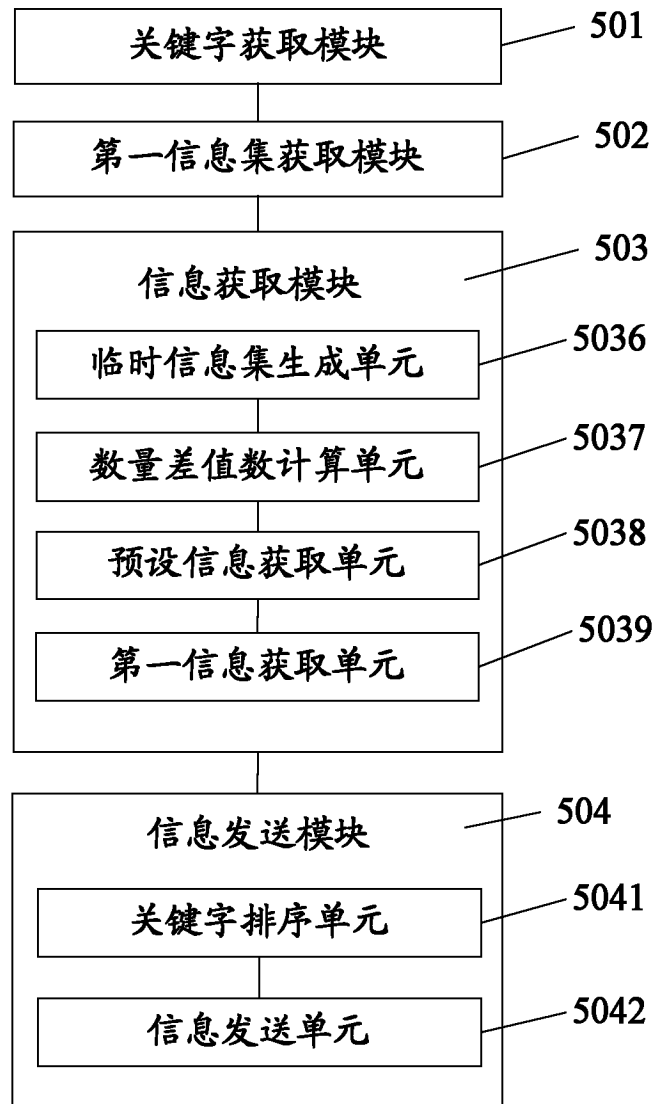


图 7

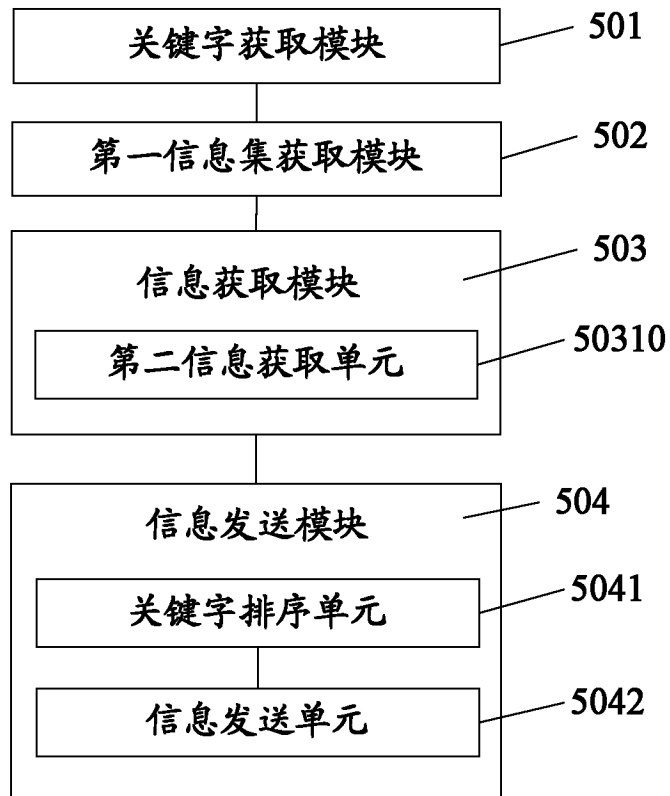


图 8