(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112115232 B
(45)授权公告日 2024.11.22

(21)申请号 202011016203.1
(22)申请日 2020.09.24
(65)同一申请的已公布的文献号
申请公布号 CN 112115232 A
(43)申请公布日 2020.12.22
(73)专利权人 腾讯科技(深圳)有限公司
地址 518057 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 韩时通
(74)专利代理机构 广州三环专利商标代理有限公司 44202
专利代理师 熊永强 杜维
(51)Int.Cl.
G06F 16/33(2019.01)
G06F 16/338(2019.01)
G06F 16/31(2019.01)
G06F 18/22(2023.01)
(56)对比文件
CN 111414763 A,2020.07.14
审查员 郭坚

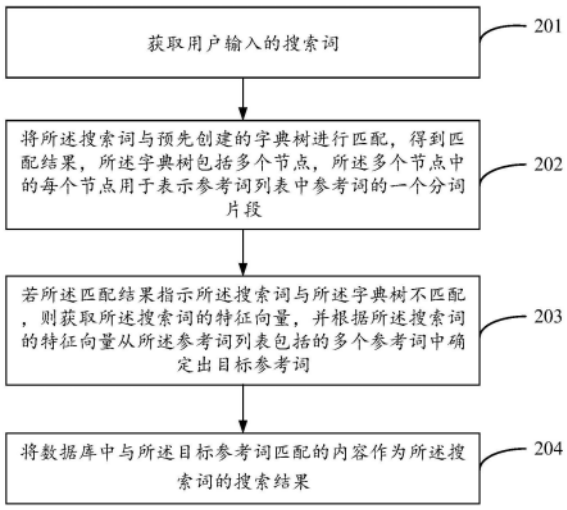
权利要求书2页 说明书13页 附图5页

(54)发明名称

一种数据纠错方法、装置及服务器

(57)摘要

本发明实施例公开了一种数据纠错方法、装置及服务器,该方法包括:获取用户输入的搜索词;将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。该方法可以准确地对搜索词进行自动化纠错,提升数据查询的效率和准确度。



CN 112115232 B

1.一种数据纠错方法,其特征在于,应用于中小型门户网站对搜索词的纠错,包括:

从数据库包括的内容中提取关键数据,所述关键数据包括多个关键词以及所述多个关键词中每个关键词的出现次数;

获取用户一定时间内的搜索记录,所述用户搜索记录包括多个搜索词以及所述多个搜索词中每个搜索词的出现次数;

根据所述关键数据和所述用户搜索记录创建参考词列表;

获取用户输入的搜索词;

将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;

若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果;

获取搜索词纠错日志,所述搜索词纠错日志包括多个纠错记录,所述多个纠错记录中的每个纠错记录包括输入的搜索词以及对应的目标参考词;

获取根据所述搜索词纠错日志中出错的纠错记录输入的待添加参考词;

将所述待添加参考词添加到所述参考词列表中,并更新所述字典树。

2.根据权利要求1所述的方法,其特征在于,所述将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,包括:

对所述搜索词进行分词处理,得到所述搜索词的多个分词片段;

将所述多个分词片段中的每个分词片段与预先创建的字典树中的各个节点进行匹配;

若存在分词片段与所述字典树中的节点不匹配,则生成匹配结果,所述匹配结果用于指示所述搜索词与所述字典树不匹配。

3.根据权利要求1所述的方法,其特征在于,所述根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词,包括:

获取所述参考词列表包括的多个参考词中每个参考词的特征向量;

计算所述搜索词的特征向量与所述每个参考词的特征向量之间的相似度;

将对应的相似度最高的参考词作为目标参考词。

4.根据权利要求3所述的方法,其特征在于,所述参考词列表还包括所述多个参考词中每个参考词的词频,所述将对应的相似度最高的参考词作为目标参考词,包括:

获取对应的相似度最高的第一参考词和对应的相似度次高的第二参考词;

获取所述第一参考词对应的相似度和所述第二参考词对应的相似度之间的差值;

判断所述差值是否小于或等于预设差值阈值;

若是,则从所述参考词列表中查询所述第一参考词的词频和所述第二参考词的词频,并将所述第一参考词和所述第二参考词中词频最高的参考词作为目标参考词;

若否,则将所述第一参考词作为目标参考词。

5.根据权利要求1所述的方法,其特征在于,所述方法还包括:

若所述匹配结果指示所述搜索词与所述字典树匹配,则从数据库中查询与所述搜索词匹配的候选内容;

获取所述候选内容与所述搜索词之间的相关度;

若所述相关度小于或等于预设相关度阈值,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

将所述数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

6.根据权利要求1~5中任一项所述的方法,其特征在于,所述将所述搜索词与预先创建的字典树进行匹配,得到匹配结果之前,所述方法还包括:

获取参考词列表,所述参考词列表包括多个参考词;

对所述多个参考词中的每个参考词进行分词处理,得到所述每个参考词的多个分词片段;

根据所述每个参考词的每个分词片段生成字典树的节点,以创建所述多个参考词对应的字典树。

7.一种数据纠错装置,其特征在于,应用于中小型门户网站对搜索词的纠错,包括:

数据获取模块,用于从数据库包括的内容中提取关键数据,所述关键数据包括多个关键词以及所述多个关键词中每个关键词的出现次数;获取用户一定时间内的搜索记录,所述用户搜索记录包括多个搜索词以及所述多个搜索词中每个搜索词的出现次数;根据所述关键数据和所述用户搜索记录创建参考词列表;获取用户输入的搜索词;

数据匹配模块,用于将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;

数据确定模块,用于若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

数据输出模块,用于将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果;

所述数据获取模块,还用于获取搜索词纠错日志,所述搜索词纠错日志包括多个纠错记录,所述多个纠错记录中的每个纠错记录包括输入的搜索词以及对应的目标参考词;获取根据所述搜索词纠错日志中出错的纠错记录输入的待添加参考词;将所述待添加参考词添加到所述参考词列表中,并更新所述字典树。

8.一种服务器,其特征在于,包括存储器以及处理器,所述存储器存储一组程序代码,所述处理器调用所述存储器中存储的程序代码,用于执行权利要求1~6任一项所述的方法。

一种数据纠错方法、装置及服务器

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种数据纠错方法、装置及服务器。

背景技术

[0002] 随着互联网技术的飞速发展,互联网中的信息量也越来越大,如何能够更有效地获取其中所需的信息,已经越来越受到人们的关注。大多数人是通过搜索引擎来完成他们信息的搜寻过程,但是当用户在搜索引擎中输入搜索词进行查询时,往往出于各种原因,总会存在输入错别字、多字或少字的情况,例如,用户在存在着同音别字的情况时,将“公积金”输入成“公鸡金”,搜索引擎可能会发生返回的搜索结果不符合用户预期的问题,此时用户需要在大量的搜索结果页面寻找所需信息,通常需要花费较多时间查阅搜索结果后发现搜索词输入错误,并尝试更正搜索词重新搜索,或者为了得到有效信息而不停地更换搜索词,这种搜索方法无法达到智能化地查询的目的,并且效率较低。

发明内容

[0003] 有鉴于此,本发明实施例提供了一种数据纠错方法,可以准确地对搜索词进行自动化纠错,提升数据查询的效率和准确度。

[0004] 第一方面,本发明实施例提供了一种数据纠错方法,包括:

[0005] 获取用户输入的搜索词;

[0006] 将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;

[0007] 若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

[0008] 将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0009] 第二方面,本发明实施例提供了一种数据纠错装置,该装置包括:

[0010] 数据获取模块,用于获取用户输入的搜索词;

[0011] 数据匹配模块,用于将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;

[0012] 数据确定模块,用于若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

[0013] 数据输出模块,用于将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0014] 第三方面,本申请实施例提供了一种服务器,该设备包括处理器、输入设备、输出设备和存储器,所述处理器、输入设备、输出设备和存储器相互连接,其中,所述存储器用于

存储计算机程序,所述计算机程序包括程序指令,所述处理器被配置用于调用所述程序指令,用于执行上述一种数据纠错方法所涉及到的操作。

[0015] 第四方面,本发明实施例提供一种计算机可读存储介质,存储有计算机程序,所述处理器执行上述一种数据纠错方法所涉及的程序。

[0016] 第五方面,本申请实施例提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述一种数据纠错方法。

[0017] 本发明实施例对于获得的搜索词,首先将搜索词与预先创建的字典树进行匹配,根据得到的匹配结果来确定搜索词是否需要纠错,当匹配结果指示搜索词与字典树不匹配时,获取搜索词的特征向量,并根据搜索词的特征向量与参考词列表包括的多个参考词之间的相似度来确定出目标参考词,最后将数据库中与目标参考词匹配的内容作为搜索词的搜索结果,可以准确地对搜索词进行自动化纠错,提升数据查询的效率和准确度。

附图说明

[0018] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0019] 图1是本发明实施例提供的一种数据检索系统的架构示意图;

[0020] 图2是本发明实施例提供的一种数据纠错方法的流程示意图;

[0021] 图3是本发明实施例提供的创建字典树的步骤的流程示意图;

[0022] 图4是本发明实施例提供的一种字典树的结构示意图;

[0023] 图5是本发明实施例提供的纠错日志界面示意图;

[0024] 图6是本发明实施例提供的一种数据纠错装置的结构示意图;

[0025] 图7是本发明实施例提供的一种服务器的结构示意图。

具体实施方式

[0026] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0027] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、云存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0028] 云计算(cloud computing)是一种计算模式,它将计算任务分布在大量计算机构成的资源池上,使各种应用系统能够根据需要获取计算力、存储空间和信息服务。提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的,并且可以随时获取,按

需使用,随时扩展,按使用付费。

[0029] 云存储(cloud storage)是在云计算概念上延伸和发展出来的一个新的概念,分布式云存储系统指通过集群应用、网格技术以及分布存储文件系统等功能,将网络中大量各种不同类型的存储设备(存储设备也称之为存储节点)通过应用软件或应用接口集合起来协同工作,共同对外提供数据存储和业务访问功能的一个存储系统。

[0030] 数据库(Database),简而言之可视为电子化的文件柜——存储电子文件的处所,用户可以对文件中的数据进行新增、查询、更新、删除等操作。所谓"数据库"是以一定方式储存在一起、能与多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立的数据集合。

[0031] 本申请提供的数据纠错方法对输入的搜索词进行纠错时,需要涉及人工智能技术中的云计算、云存储、数据库等技术,通过将用户输入的搜索词与字典树进行匹配,能够实现搜索词的自动化纠错,提升数据查询的效率和准确度。

[0032] 在对本申请实施例进行详细地解释说明之前,先对本申请实施例的应用场景予以说明。

[0033] 本申请实施例中的数据纠错方法具体可以应用于一些中小型门户网站对搜索词进行纠错,比如,政务服务网站,目前政务服务网站存在技术积累少,缺乏网络相关人才,以至于很难对网站进行后续的运营,同时在系统上线后只保证可用性,不保证实用性的问题,因此当用户在输入错误的搜索词时,很大程度上会导致返回的搜索结果不满足用户的需求,需要人为地发现到搜索词出现错误后,尝试更正搜索词重新搜索。政务服务网站只是用来示例,还可以应用于其他中小型门户网站,如企业门户网站。

[0034] 如图1所示,是本发明实施例提供的一种数据检索系统的架构示意图。该数据检索系统可以包括用户终端101、网络102和服务器103,用户终端101和服务器103通过网络102进行通信。用户终端101获取搜索词,该搜索词可以是基于用户终端101的用户输入的确定的文本,再通过网络102将搜索词发送至服务器103,服务器103对该搜索词进行匹配,确定直接使用搜索词能否得到准确的搜索结果,例如判断搜索词是否在字典树中,如果搜索词不在字典树中判定不匹配,则对搜索词自动进行纠错,并将纠错后得到的目标参考词对应的匹配内容作为搜索词的搜索结果。其中,网络102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,用户终端101可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,服务器103可以用独立的服务器或者是多个服务器组成的服务器集群来实现,例如政务平台的政务服务器或者服务器集群,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0035] 可以理解的是,本申请实施例描述的系统的架构示意图是为了更加清楚的说明本申请实施例的技术方案,并不构成对于本申请实施例提供的技术方案的限定,本领域普通技术人员可知,随着系统架构的演变和新业务场景的出现,本申请实施例提供的技术方案对于类似的技术问题,同样适用。

[0036] 在一个实施例中,如图2所示,是本发明实施例基于图1的数据检索系统提供的一种数据纠错方法。本实施例主要以该方法应用于上述图1中的服务器103来举例说明,包括以下步骤:

[0037] 步骤S201、获取用户输入的搜索词。

[0038] 本发明实施例中,用户需要通过用户终端进行信息搜索时,可以在搜索框中输入搜索词,以使得用户终端获取到用户输入的搜索词,用户终端将该搜索词发送给服务器,服务器获取到该搜索词。其中搜索框是指搜索引擎系统中的交互控件,用于根据在搜索框中输入的搜索字符提取海量信息中相应的准确内容。

[0039] 需要说明的是,在实际应用中,用户在搜索框中输入搜索词时,可以手动输入搜索词,也可以采用语音的形式输入搜索词,等等,本申请实施例中对用户输入搜索词的方式不做限制。

[0040] 步骤S202、将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段。

[0041] 其中,字典树(Trie树)又被称为前缀树,是一种树状的数据结构,包括多个节点,可用于字符串匹配和快速查找等处理过程中。它可以最大限度地减少无谓字符串的比较次数,提高词频统计和字符串排序的效率。其核心思想是通过构建树状结构,用空间换时间,利用字符串间的公共前缀来降低查询的开销。字典树一般具有如下三个特质:1)根节点不包含字符,除根节点外每个节点只包含一个字符串,该字符串具体可以是参考词列表中参考词的一个分词片段;2)从根节点到某一叶子节点,路径上所有字符串连起来,就是该节点对应的组合字符串,每个组合字符串具体可以是一个参考词;3)每个节点的所有子节点包含的字符都不相同。

[0042] 在本申请实施例中,预先创建的字典树通过获取参考词列表中的多个参考词,然后对于已经获得的多个参考词,对每个参考词进行分词处理,得到每个参考词的多个分词片段,以每个分词片段为单位,将参考词列表中的每个参考词的分词片段依次存储至字典树的一条路径中的不同节点中,在创建字典树的过程中,要对比参考词的第一个(或出现较早的)分词片段的字符是否已存在节点,存在则把指针指向该节点,无则为该分词片段创建节点。

[0043] 在一个实施例中,服务器从数据库中获取大量的原始语料,这些语料中包含有一些关键数据,关键数据包括关键词和对应的出现次数。服务器也可以从搜索引擎中获取用户在一定时间内的搜索记录,并获取到多个搜索词和对应的出现次数,通过统计关键数据和用户搜索记录中的关键词和搜索词,生成参考词列表。

[0044] 具体地,服务器在接收到用户终端输入的搜索词后,则对搜索词进行分词处理,得到搜索词的多个分词片段,从搜索词的第一个分词片段开始,与预先创建的字典树从根节点的第一层子节点开始,匹配当前分词片段是否已存在节点,当存在节点与分词片段匹配时,则将搜索词的下一个分词片段从字典树中的当前节点继续进行匹配,当存在搜索词中的分词片段不在字典树中的节点时,则生成匹配结果,认为搜索词与字典树不匹配。

[0045] 作为本实施例的一个具体示例,若用户终端输入的搜索词为“驾照申令”,字典树包括的是参考词“驾照申领”的各个分词片段对应的节点,对搜索词进行分词处理后,得到的“驾”“照”“申”“令”即为搜索词的多个分词片段,将搜索词的多个分词片段“驾”“照”“申”“令”与字典树中的节点进行匹配,当匹配到分词片段“令”时,由于将“领”输入成“令”导致无法在字典树中找到相应的节点,因此判定搜索词与字典树不匹配。

[0046] 步骤S203、若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词。

[0047] 其中,特征向量是指将自然语言表示成能够表达自身语义的向量,可以使用文本表示算法将搜索词和参考词列表中的多个参考词进行向量化,得到相应的特征向量,文本表示算法包括基于向量空闲模型的方法、基于主题模型的方法、基于神经网络的方法等等。

[0048] 具体地,为了确定目标参考词,首先需要获得参考词列表中的多个参考词中每个参考词的特征向量,在获取到参考词列表包括的多个参考词中每个参考词的特征向量后,计算搜索词的特征向量与每个参考词的特征向量之间的相似度,在对相似度进行排序后,将最高相似度对应的参考词作为目标参考词。

[0049] 其中,相似度用于衡量搜索词的特征向量与参考词列表中的多个参考词中每个参考词的特征向量的相似性,可以使用相似度算法计算相似度,相似度算法包括但不限于欧几里得距离算法、余弦相似度算法、皮尔逊相关系数算法、杰卡德相似系数算法等等。

[0050] 例如,用户终端输入的搜索词为“驾照申令”,参考词列表中包括“驾照申领”这一参考词。服务器在将“驾照申令”转换为特征向量后可以与参考词列表中的多个参考词中每个参考词的特征向量进行相似度计算,当返回“驾照申领”为相似度最高的参考词时,将“驾照申领”作为目标搜索词。

[0051] 步骤S204、将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0052] 具体地,服务器在获取到目标搜索词后,直接根据目标搜索词在数据库中查找,并将查找到的相应的搜索结果作为搜索词的搜索结果。例如,当用户终端输入的搜索词为“驾照申令”时,通过步骤S202和步骤S203获得目标搜索词“驾照申领”后,直接用“驾照申领”作为目标参考词在数据库中搜索相应的数据资源,作为搜索词的搜索结果。

[0053] 上述基于数据纠错的方法,在获取到搜索词后,根据搜索词与预先创建的字典树进行匹配,根据得到的匹配结果来确定搜索词是否需要纠错,当匹配结果指示搜索词与字典树不匹配时,根据搜索词的特征向量与参考词列表包括的多个参考词之间的相似度来确定出目标参考词,最后将数据库中与目标参考词匹配的内容作为搜索词的搜索结果,能够智能化地判断搜索词是否存在错误并进行纠错,不需要用户再去手动更正搜索词,能够有效提高查询数据的效率。

[0054] 在一个实施例中,如图3所示,将所述搜索词与预先创建的字典树进行匹配,得到匹配结果之前,还包括创建字典树的步骤,该步骤具体包括以下内容:

[0055] 步骤301、从数据库包括的内容中提取关键数据,所述关键数据包括多个关键词以及所述多个关键词中每个关键词的出现次数;

[0056] 服务器从数据库中获取大量的原始语料,这些语料中包含有一些关键词(比如“驾照申领”等),并统计出这些关键词的出现次数。

[0057] 步骤302、获取用户搜索记录,所述用户搜索记录包括多个搜索词以及所述多个搜索词中每个搜索词的出现次数;

[0058] 具体地,服务器可以从搜索引擎中获取用户在一定时间内的搜索记录,并获取到多个搜索词和对应的出现次数。

[0059] 在一个实施例中,服务器从搜索引擎中获取用户在一定时间内的搜索记录,将搜索记录按照时间排序,对于某一个用户的搜索记录来说,通常用户的搜索行为是分段的,每段之间有较为明显的间隔,将每段称为一个搜索session,服务器获取用户每一个搜索session的多个搜索词,并统计对应的出现次数,由于用户在一个搜索session内的搜索行为都是为了解决一个问题,因此一个搜索session内用户输入的搜索词往往都是相关的,通过统计用户在一个搜索session的搜索词,可以挖掘出重复的词或同义词,进一步提升参考词列表的完备性和提高查询的准确度。

[0060] 在一个实施例中,也可以先执行步骤302,再执行步骤301。本发明实施例对步骤301和302的具体顺序不做限定。

[0061] 步骤303、根据所述关键数据和所述用户搜索记录创建参考词列表。

[0062] 具体地,通过统计关键数据和用户搜索记录中的关键词和搜索词,生成参考词列表中的参考词列,并将参考词的出现次数作为参考词对应的词频。

[0063] 示例性地,参考词列表如表1所示:

[0064] 表1参考词列表

[0065]

参考词	词频	创建时间
护士长	42	2020-09-18 10:04:01
公积金	210	2020-09-18 10:04:01
生育津贴	21	2020-09-18 10:04:01
生育补贴	22	2020-09-18 10:04:01
驾照申领	125	2020-09-18 10:04:01
西山居	1	2020-09-18 10:04:01

[0066] 步骤304、获取参考词列表,所述参考词列表包括多个参考词;

[0067] 步骤305、对所述多个参考词中的每个参考词进行分词处理,得到所述每个参考词的多个分词片段;

[0068] 本实施例中,分词处理是指将文本序列切分成一个一个单独的字,分词片段是指对文本序列进行处理后得到的字片段。例如,用户输入的搜索词为“驾照申领”,利用分词处理后得到“驾”“照”“申”“领”,其中的“驾”“照”“申”“领”分别为一个分词片段。

[0069] 步骤306、根据所述每个参考词的每个分词片段生成字典树的节点,以创建所述多个参考词对应的字典树。

[0070] 在本实施例中,将参考词列表中的每个参考词的分词片段依次存储至字典树的一条路径中的不同节点中,在创建字典树的过程中,要对比参考词的第一个(或出现较早的)分词片段的字符,是否已存在节点,存在则把指针指向该节点,无则为该分词片段创建节点。

[0071] 举例说明,如图4所示,以参考词列表包括的“公积金”、“生育津贴”、“生育补贴”这三个参考词为例,服务器可以创建一个根节点,根节点下面的多个节点均为子节点,针对参考词“公积金”、“生育津贴”、“生育补贴”中的分词片段“公”和“生”,服务器可以确定不存在与根节点相连的、且与该字符匹配的子节点,服务器即创建一个与根节点相连的子节点“公”和“生”。类似地,针对参考词“公积金”中的分词片段“积”,服务器可以确定不存在与节点“公”相连的、且与该字符匹配的孩子节点,然后,服务器可以创建一个与节点“公”相连的

孩子节点“积”;针对参考词“公积金”中的分词片段“金”,服务器可以确定不存在与节点“积”相连的、且与该字符匹配的孩子节点,然后,服务器可以创建一个与节点“积”相连的孩子节点“金”;由此服务器可以得到如图4中(1)所示的树形结构。类似地,可以得到如图4所示的树形结构。

[0072] 在一个实施例中,将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,包括:对所述搜索词进行分词处理,得到所述搜索词的多个分词片段;将所述多个分词片段中的每个分词片段与预先创建的字典树中的各个节点进行匹配;若存在分词片段与所述字典树中的节点不匹配,则生成匹配结果,所述匹配结果用于指示所述搜索词与所述字典树不匹配。

[0073] 在本实施例中,服务器在接收到用户终端输入的搜索词后,则对搜索词进行分词处理,得到搜索词的多个分词片段,从搜索词的第一个分词片段开始,与预先创建的字典树从根节点的第一层子节点开始,匹配当前分词片段是否已存在节点,当存在节点与分词片段匹配时,则将搜索词的下一个分词片段从字典树中的当前节点继续进行匹配,当存在搜索词中的分词片段不在字典树中的节点时,则生成匹配结果,认为搜索词与字典树不匹配。

[0074] 作为本实施例的一个具体示例,若用户终端输入的搜索词为“驾照申令”,对搜索词进行分词处理后,得到的“驾”“照”“申”“令”则为搜索词的多个分词片段,将搜索词的多个分词片段“驾”“照”“申”“令”与字典树中的节点进行匹配,当匹配到分词片段“令”时,由于将“领”输入成“令”导致无法在字典树中找到相应的节点,因此判定搜索词与字典树不匹配。

[0075] 上述实施例根据获得的搜索词,与字典树进行匹配,能够高效地确认搜索词是否需要纠错,以保证查询返回的搜索结果的准确性。

[0076] 在一个实施例中,服务器在获取到参考词列表包括的多个参考词中每个参考词的特征向量后,计算搜索词的特征向量与每个参考词的特征向量之间的相似度时,可以通过预置的多个线程对参考词列表中的各个子参考词列表进行并发计算,得到各个子参考词列表的最高相似度对应的参考词。具体地,服务器可以使用bigram、trigram等策略来控制拆分大小,从而控制计算量,将参考词列表分成K个子参考词列表,其中K大于等于2,每个子参考词列表包含N个参考词,其中N大于等于1;服务器调用预置的多个线程分别计算用户输入的搜索词的特征向量与每个子参考词列表中的N个参考词的特征向量的相似度,得到每个子参考词列表中的最高相似度对应的参考词,并根据相似度降序排序,得到最终的最高相似度对应的目标搜索词。

[0077] 需要说明的是,在本实施例中,每个线程可以计算搜索词的特征向量与一个子参考列表中的N个参考词的特征向量的相似度,多个线程之间可以同时进行处理,不会互相影响,从而能够降低搜索所需的时间,提高搜索的处理速度。

[0078] 作为本实施例的一个具体示例,如表2所示,服务器将表1参考词列表拆分成两个参考词列表,并获取每个参考词列表中的多个参考词对应的特征向量,得到参考词特征向量列表(a)和参考词特征向量列表(b),例如,用户终端输入搜索词“驾驶证申领”,服务器在将“驾驶证申领”转换为特征向量后可以与参考词特征向量列表(a)和参考词特征向量列表(b)中每个参考词的特征向量进行相似度计算,其中与参考词特征向量列表(a)进行计算后返回[公积金=0.0123],与参考词特征向量列表(b)进行计算后返回[驾驶证申领=

0.89102],服务器则从参考词特征向量列表(a)和(b)返回的结果中选择相似度最高的参考词“驾驶证申领”作为目标搜索词。

[0079] 示例性地,参考词特征向量列表如表2所示:

[0080] 表2参考词特征向量列表(a)

[0081]

参考词	特征向量
护士证	[0.1233422,10.1292920d,101.929101]
公积金	[1.1233422,10.1292920d,101.929101]
生育津贴	[2.1233422,10.1292920d,101.929101]

[0082] 表2参考词特征向量列表(b)

[0083]

参考词	特征向量
生育补贴	[3.1233422,10.1292920d,101.929101]
驾照申领	[4.1233422,10.1292920d,101.929101]
西山居	[5.1233422,10.1292920d,101.929101]

[0084] 在其中一个实施例中,在上述实施例的基础上,在获得参考词列表后,还可以将参考词列表中的参考词转化为拼音,将获得的参考词的拼音进行分词处理,并将进行分词处理后的拼音的分词片段构建拼音字典树。

[0085] 示例性地,参考词的拼音列表如表3所示:

[0086] 表3参考词的拼音列表

[0087]

参考词	拼音	创建时间
护士长	hushizhang	2020-09-18 10:05:10
公积金	gongjijin	2020-09-18 10:05:10
生育津贴	shengyujintie	2020-09-18 10:05:10
生育补贴	shengyubutie	2020-09-18 10:05:10
驾照申领	jiazhaoshenling	2020-09-18 10:05:10
西山居	xishanju	2020-09-18 10:05:10

[0088] 服务器在将用户输入的搜索词与字典树进行匹配后,匹配结果显示搜索词的分词片段不能在字典树中找到对应的节点,进一步地,将搜索词对应的拼音与拼音字典树进行匹配,若搜索词对应的拼音能在拼音字典树中找到相应的节点,则将匹配节点的拼音字符序列对应的参考词作为目标参考词;若搜索词对应的拼音不能在拼音字典树中找到相应的节点,将搜索词对应的拼音与参考词的拼音列表包括的多个拼音中每个拼音的特征向量进行相似度计算,获得对应的最高相似度的拼音,将这个拼音对应的参考词作为目标参考词。

[0089] 在一个实施例中,参考词列表还包括所述多个参考词中每个参考词的词频,所述将对应的相似度最高的参考词作为目标参考词,包括:获取对应的相似度最高的第一参考词和对应的相似度次高的第二参考词;获取所述第一参考词对应的相似度和所述第二参考词对应的相似度之间的差值;判断所述差值是否小于或等于预设差值阈值;若是,则从所述参考词列表中查询所述第一参考词的词频和所述第二参考词的词频,并将所述第一参考词和所述第二参考词中词频最高的参考词作为目标参考词;若否,则将所述第一参考词作为目标参考词。

[0090] 在本实施例中,将参考词的特征向量与参考词列表中的多个参考词的特征向量进行相似度计算后,获取对应的相似度最高的第一参考词和对应的相似度次高的第二参考词,当第一参考词和第二参考词与搜索词之间的编辑距离都比较小时,第一参考词和第二参考词对应的相似度之间的差值会小于或等于预设差值阈值,将第一参考词和第二参考词对应的词频作为依据,选取词频最高的参考词作为目标参考词。例如,当用户输入的搜索词为“生育”时,参考词列表中包括参考词“生育津贴”和“生育补贴”,搜索词“生育”与参考词“生育津贴”和“生育补贴”的编辑距离都为2,此时搜索词“生育”与参考词“生育津贴”和“生育补贴”对应的相似度之间的差值小于预设差值阈值,由于参考词“生育津贴”的词频为21和参考词“生育补贴”的词频为22,因此将“生育补贴”作为目标参考词。

[0091] 在一个实施例中,数据纠错方法还包括以下步骤:若所述匹配结果指示所述搜索词与所述字典树匹配,则从数据库中查询与所述搜索词匹配的候选内容;获取所述候选内容与所述搜索词之间的相关度;若所述相关度小于或等于预设相关度阈值,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;将所述数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0092] 在本实施例中,将用户输入的搜索词的分词片段与字典树进行匹配,用户输入的搜索词的分词片段在字典树中都能找到相应的节点时,匹配结果指示搜索词与字典树匹配,将用户输入的搜索词在数据库中查询,获得相应的搜索结果作为候选内容。

[0093] 进一步地,获取搜索词与候选内容之间的相关度,例如在数据库中返回的候选内容具有十篇网页文章,利用相关度算法,例如BM25算法,计算搜索词与每一篇网页文章的相关度,最终将搜索词与十篇网页文章的相关度的均值与预设相关度阈值进行比较,假设100分为满分,预设相关度阈值为50分,则用户输入的搜索词与候选内容的相关度低于50分的话,则也会根据用户输入的搜索词的特征向量从参考词列表包括的多个参考词中确定出目标参考词。

[0094] 在一个实施例中,将用户输入的搜索词在数据库中查询,获得相应的搜索结果作为候选内容后,还可以利用点击模型计算候选内容是否是用户感兴趣的搜索结果,例如计算用户点击候选内容中的网页文章的点击概率,当预测的点击概率低于预设阈值时,根据搜索词的特征向量从参考词列表包括的多个参考词中确定出目标参考词,并将数据库中与目标参考词匹配的内容作为搜索词的搜索结果,通过利用点击模型预先预测用户对候选内容的点击概率来自动替换输入的搜索词,可以避免用户花费时间在不感兴趣的搜索结果上,提高用户查询的效率。

[0095] 其中,点击模型通过挖掘搜索词、搜索词对应的搜索内容和搜索词对应的搜索内容中被点击的搜索结果等信息,基于一些前提假设建立概率图模型,从而对用户的搜索行为进行建模。点击模型包括但不限于级联模型、动态贝叶斯网络模型等等。

[0096] 在一个实施例中,数据纠错方法还包括以下步骤:获取搜索词纠错日志,所述搜索词纠错日志包括多个纠错记录,所述多个纠错记录中的每个纠错记录包括输入的搜索词以及对应的目标参考词;获取根据所述搜索词纠错日志中出错的纠错记录输入的待添加参考词;将所述待添加参考词添加到所述参考词列表中,并更新所述字典树。

[0097] 如图5所示,通过纠错日志对用户输入的搜索词和对应的目标搜索词进行记录,得

到多个纠错记录。在政务网站里,实际上是没有“驾照”这一词的,正确的是驾驶证,因此不在表1参考词列表中添加驾驶证之前,很有可能会被纠错为护照。当运营人员从服务器中获取纠错日志后发现搜索自动纠错错误,可以将“驾驶证”作为待添加参考词,服务器将待添加参考词“驾驶证”添加到参考词列表。

[0098] 进一步地,服务器在字典树中查找该待添加参考词,将字典树中的节点与待添加参考词的分词片段进行匹配,当匹配结果指示该待添加参考词不能在字典树中找到相应的节点进行匹配时,则对字典树的节点进行更新。由此能有效地根据纠错日志对字典树中的节点进行动态更新,从而有效扩充了参考词,增强了字典树中存储的参考词的完备性和准确度。

[0099] 如图6所示,图6是本申请实施例提供的一种数据纠错装置的结构示意图,包括:

[0100] 数据获取模块601,用于获取用户输入的搜索词;

[0101] 数据匹配模块602,用于将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;

[0102] 数据确定模块603,用于若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

[0103] 数据输出模块604,用于将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0104] 在一个实施例中,数据匹配模块602将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,包括:

[0105] 对所述搜索词进行分词处理,得到所述搜索词的多个分词片段;

[0106] 将所述多个分词片段中的每个分词片段与预先创建的字典树中的各个节点进行匹配;

[0107] 若存在分词片段与所述字典树中的节点不匹配,则生成匹配结果,所述匹配结果用于指示所述搜索词与所述字典树不匹配。

[0108] 在一个实施例中,数据确定模块603根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词,包括:

[0109] 获取所述参考词列表包括的多个参考词中每个参考词的特征向量;

[0110] 计算所述搜索词的特征向量与所述每个参考词的特征向量之间的相似度;

[0111] 将对应的相似度最高的参考词作为目标参考词。

[0112] 在一个实施例中,数据确定模块603将对应的相似度最高的参考词作为目标参考词,包括:

[0113] 获取对应的相似度最高的第一参考词和对应的相似度次高的第二参考词;

[0114] 获取所述第一参考词对应的相似度和所述第二参考词对应的相似度之间的差值;

[0115] 判断所述差值是否小于或等于预设差值阈值;

[0116] 若是,则从所述参考词列表中查询所述第一参考词的词频和所述第二参考词的词频,并将所述第一参考词和所述第二参考词中词频最高的参考词作为目标参考词;

[0117] 若否,则将所述第一参考词作为目标参考词。

[0118] 在一个实施例中,若所述匹配结果指示所述搜索词与所述字典树匹配,数据确定模块603,还用于从数据库中查询与所述搜索词匹配的候选内容;

[0119] 数据确定模块603,还用于获取所述候选内容与所述搜索词之间的相关度;

[0120] 数据确定模块603,还用于若所述相关度小于或等于预设相关度阈值,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

[0121] 数据输出模块604,还用于将所述数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0122] 在一个实施例中,将所述搜索词与预先创建的字典树进行匹配,得到匹配结果之前,数据获取模块601,还用于获取参考词列表,所述参考词列表包括多个参考词;

[0123] 数据获取模块601,还用于对所述多个参考词中的每个参考词进行分词处理,得到所述每个参考词的多个分词片段;

[0124] 数据获取模块601,还用于根据所述每个参考词的每个分词片段生成字典树的节点,以创建所述多个参考词对应的字典树。

[0125] 在一个实施例中,数据获取模块601获取参考词列表,包括:

[0126] 从数据库包括的内容中提取关键数据,所述关键数据包括多个关键词以及所述多个关键词中每个关键词的出现次数;

[0127] 获取用户搜索记录,所述用户搜索记录包括多个搜索词以及所述多个搜索词中每个搜索词的出现次数;

[0128] 根据所述关键数据和所述用户搜索记录创建参考词列表。

[0129] 在一个实施例中,数据获取模块601,还用于获取搜索词纠错日志,所述搜索词纠错日志包括多个纠错记录,所述多个纠错记录中的每个纠错记录包括输入的搜索词以及对应的目标参考词;

[0130] 数据获取模块601,还用于获取根据所述搜索词纠错日志中出错的纠错记录输入的待添加参考词;

[0131] 数据获取模块601,还用于将所述待添加参考词添加到所述参考词列表中,并更新所述字典树。

[0132] 通过本申请实施例提供的数据纠错装置,在获取到搜索词后,根据搜索词与预先创建的字典树进行匹配,根据得到的匹配结果来确定搜索词是否需要纠错,当匹配结果指示搜索词与字典树不匹配时,根据搜索词的特征向量与参考词列表包括的多个参考词之间的相似度来确定出目标参考词,最后将数据库中与目标参考词匹配的内容作为搜索词的搜索结果,能够准确地对搜索词进行自动化纠错,提升数据查询的效率和准确度。

[0133] 如图7所示,图7是本申请实施例提供的一种服务器的结构示意图,该服务器内部结构如图7所示,包括输入设备701、输出设备702、处理器703、存储器704、程序705和通信总线706,其中,输入设备701、输出设备702、处理器703,存储器704通过通信总线706完成相互间的通信。

[0134] 存储器704,用于存放程序705;

[0135] 处理器703,用于执行存储器704上所存放的程序705时,实现如下步骤:

[0136] 获取用户输入的搜索词;

[0137] 将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,所述字典树包括多个节点,所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段;

[0138] 若所述匹配结果指示所述搜索词与所述字典树不匹配,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

[0139] 将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0140] 在一个实施例中,处理器703将所述搜索词与预先创建的字典树进行匹配,得到匹配结果,包括:

[0141] 对所述搜索词进行分词处理,得到所述搜索词的多个分词片段;

[0142] 将所述多个分词片段中的每个分词片段与预先创建的字典树中的各个节点进行匹配;

[0143] 若存在分词片段与所述字典树中的节点不匹配,则生成匹配结果,所述匹配结果用于指示所述搜索词与所述字典树不匹配。

[0144] 在一个实施例中,处理器703根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词,包括:

[0145] 获取所述参考词列表包括的多个参考词中每个参考词的特征向量;

[0146] 计算所述搜索词的特征向量与所述每个参考词的特征向量之间的相似度;

[0147] 将对应的相似度最高的参考词作为目标参考词。

[0148] 在一个实施例中,处理器703将对应的相似度最高的参考词作为目标参考词,包括:

[0149] 获取对应的相似度最高的第一参考词和对应的相似度次高的第二参考词;

[0150] 获取所述第一参考词对应的相似度和所述第二参考词对应的相似度之间的差值;

[0151] 判断所述差值是否小于或等于预设差值阈值;

[0152] 若是,则从所述参考词列表中查询所述第一参考词的词频和所述第二参考词的词频,并将所述第一参考词和所述第二参考词中词频最高的参考词作为目标参考词;

[0153] 若否,则将所述第一参考词作为目标参考词。

[0154] 在一个实施例中,处理器703还用于执行以下操作:

[0155] 若所述匹配结果指示所述搜索词与所述字典树匹配,则从数据库中查询与所述搜索词匹配的候选内容;

[0156] 获取所述候选内容与所述搜索词之间的相关度;

[0157] 若所述相关度小于或等于预设相关度阈值,则获取所述搜索词的特征向量,并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词;

[0158] 将所述数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果。

[0159] 在一个实施例中,将所述搜索词与预先创建的字典树进行匹配,得到匹配结果之前,处理器703还用于执行以下操作:

[0160] 获取参考词列表,所述参考词列表包括多个参考词;

[0161] 对所述多个参考词中的每个参考词进行分词处理,得到所述每个参考词的多个分词片段;

[0162] 根据所述每个参考词的每个分词片段生成字典树的节点,以创建所述多个参考词

对应的字典树。

[0163] 在一个实施例中,处理器703获取参考词列表,包括:

[0164] 从数据库包括的内容中提取关键数据,所述关键数据包括多个关键词以及所述多个关键词中每个关键词的出现次数;

[0165] 获取用户搜索记录,所述用户搜索记录包括多个搜索词以及所述多个搜索词中每个搜索词的出现次数;

[0166] 根据所述关键数据和所述用户搜索记录创建参考词列表。

[0167] 在一个实施例中,处理器703还用于执行以下操作:

[0168] 获取搜索词纠错日志,所述搜索词纠错日志包括多个纠错记录,所述多个纠错记录中的每个纠错记录包括输入的搜索词以及对应的目标参考词;

[0169] 获取根据所述搜索词纠错日志中出错的纠错记录输入的待添加参考词;

[0170] 将所述待添加参考词添加到所述参考词列表中,并更新所述字典树。

[0171] 通过本申请实施例提供的服务器,服务器获取到搜索词后,根据搜索词与预先创建的字典树进行匹配,根据得到的匹配结果来确定搜索词是否需要纠错,当匹配结果指示搜索词与字典树不匹配时,根据搜索词的特征向量与参考词列表包括的多个参考词之间的相似度来确定出目标参考词,最后将数据库中与目标参考词匹配的内容作为搜索词的搜索结果,能够准确地对搜索词进行自动化纠错,提升数据查询的效率和准确度。

[0172] 本申请实施例还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序包括程序指令,所述程序指令被处理器执行时,可执行上述实施例中服务器所执行的步骤。

[0173] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述文件管理方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory,ROM)或随机存储记忆体(Random Access Memory,RAM)等。

[0174] 本申请实施例还提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述各方法的实施例中所执行的步骤。

[0175] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

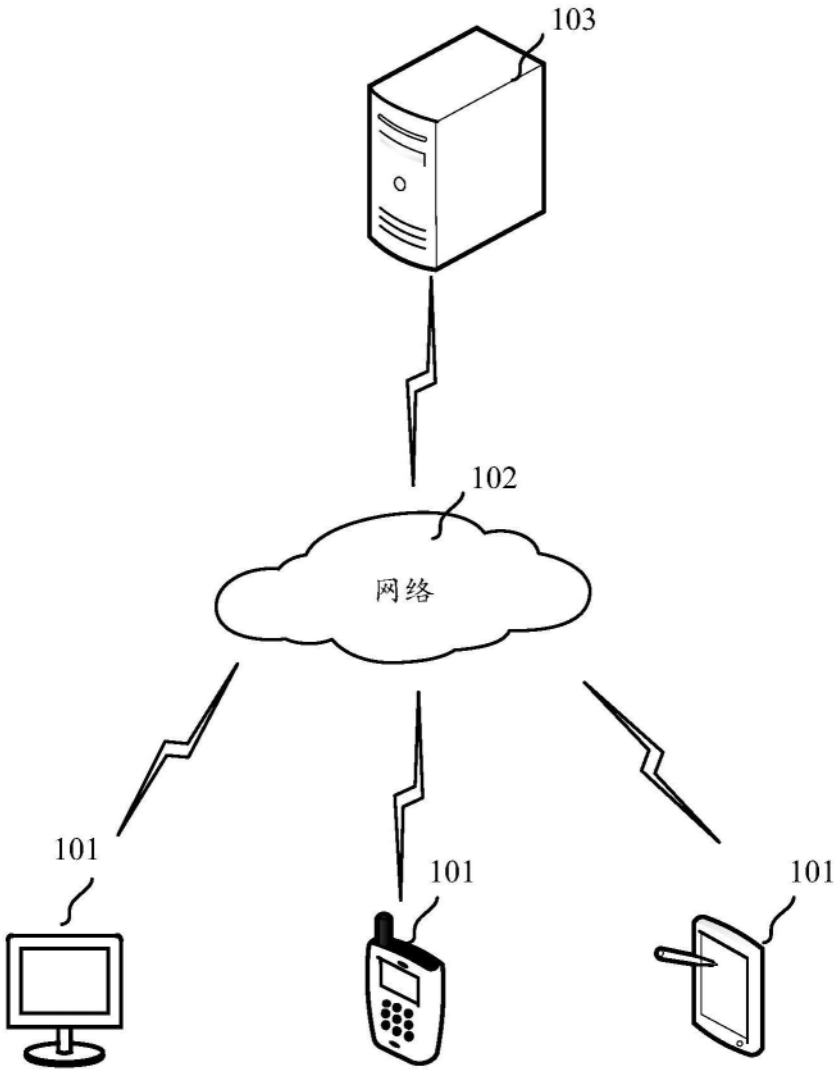


图1

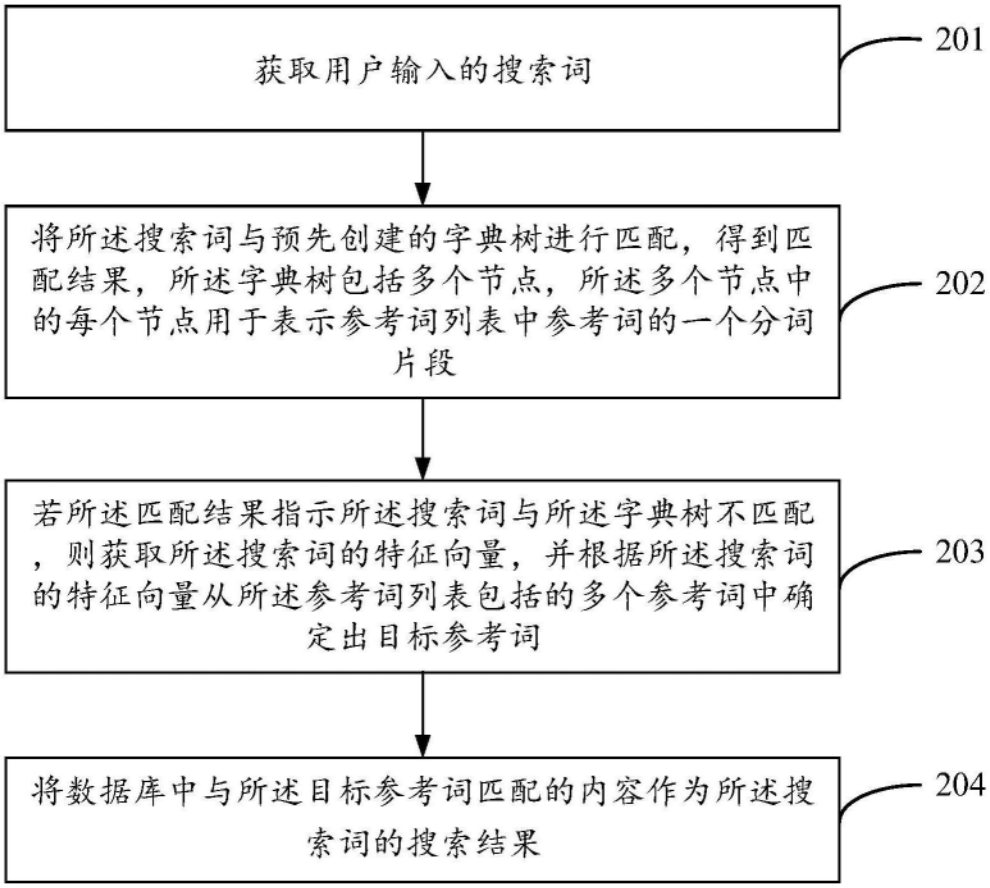

获取用户输入的搜索词
201
将所述搜索词与预先创建的字典树进行匹配，得到匹配结果，所述字典树包括多个节点，所述多个节点中的每个节点用于表示参考词列表中参考词的一个分词片段
202
若所述匹配结果指示所述搜索词与所述字典树不匹配，则获取所述搜索词的特征向量，并根据所述搜索词的特征向量从所述参考词列表包括的多个参考词中确定出目标参考词
203
将数据库中与所述目标参考词匹配的内容作为所述搜索词的搜索结果
204

图2

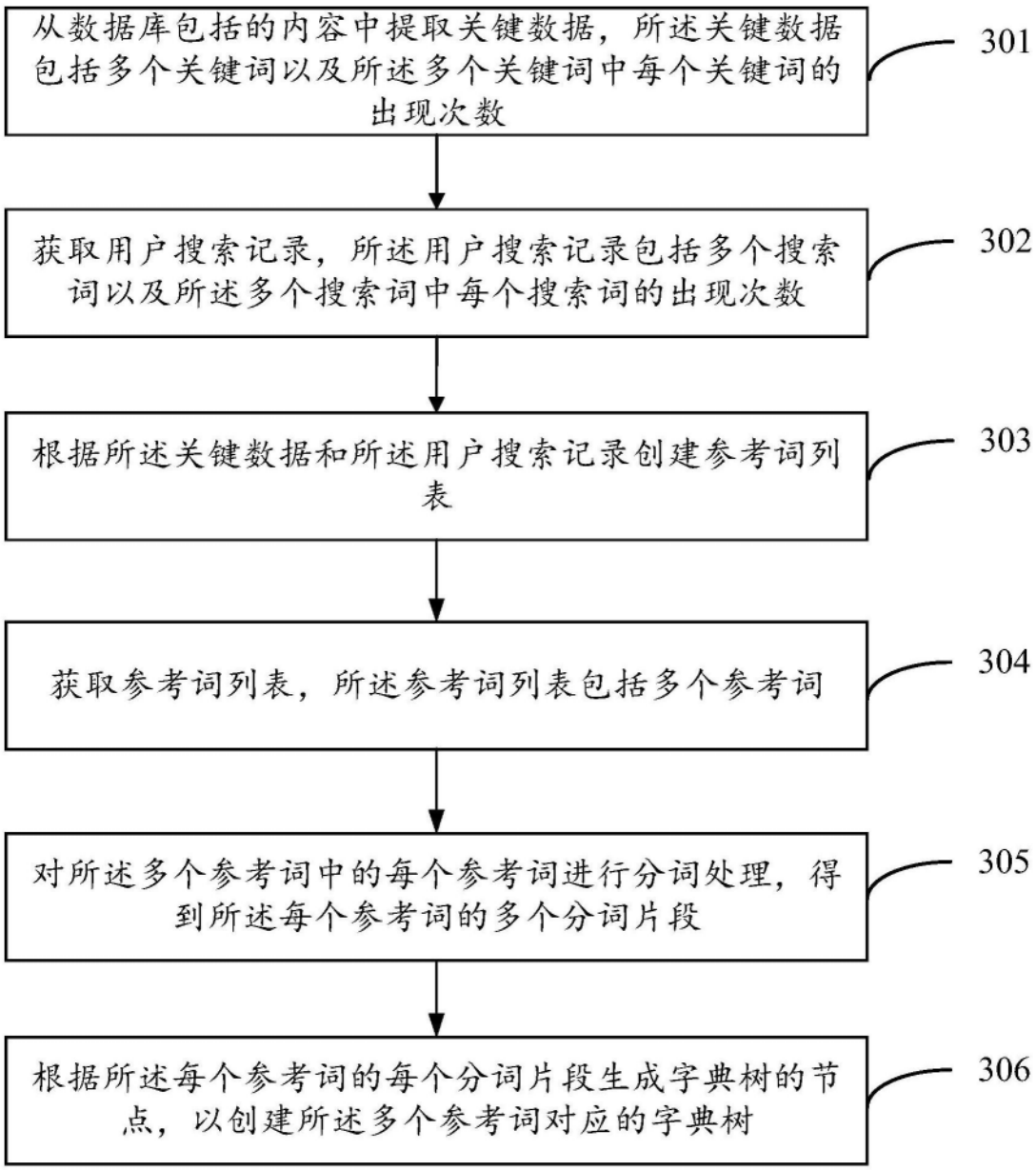

从数据库包括的内容中提取关键数据，所述关键数据包括多个关键词以及所述多个关键词中每个关键词的出现次数
301
获取用户搜索记录，所述用户搜索记录包括多个搜索词以及所述多个搜索词中每个搜索词的出现次数
302
根据所述关键数据和所述用户搜索记录创建参考词列表
303
获取参考词列表，所述参考词列表包括多个参考词
304
对所述多个参考词中的每个参考词进行分词处理，得到所述每个参考词的多个分词片段
305
根据所述每个参考词的每个分词片段生成字典树的节点，以创建所述多个参考词对应的字典树
306

图3

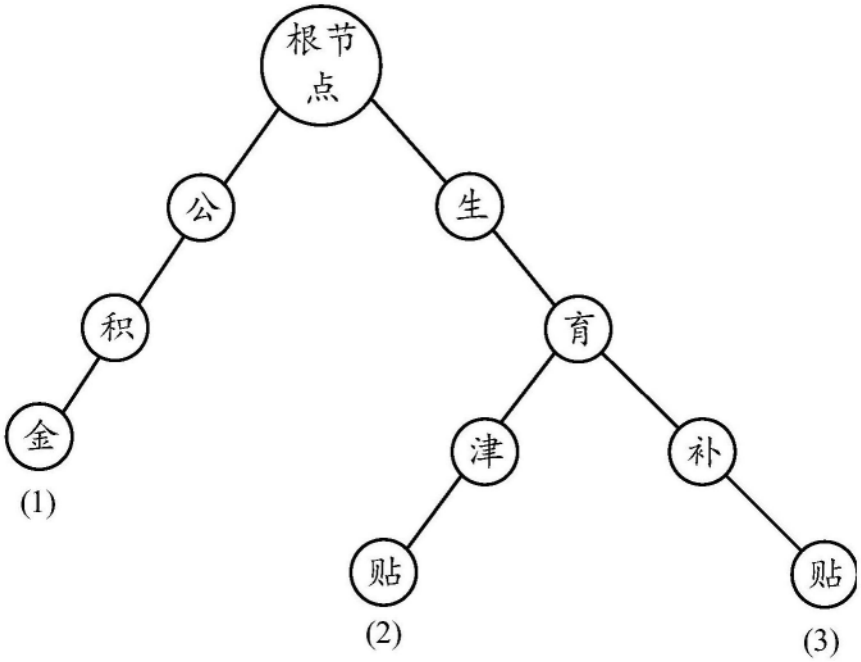

根节点
公
生
积
育
金
津
补
贴
贴
(1)
(2)
(3)

图4


搜索
请输入搜索词
创建时间
开始时间 至 结束时间
添加参考词
搜索词 是否纠错 目标参考词 创建时间
护士证 否 2020-09-19 15:30:01
公积金 是 公积金 2020-09-19 15:33:21
生育津贴 否 2020-09-19 15:41:52
生育补贴 否 2020-09-19 15:46:03
驾照申领 否 2020-09-19 15:53:01
西山居 是 西山村 2020-09-19 15:58:25

图5

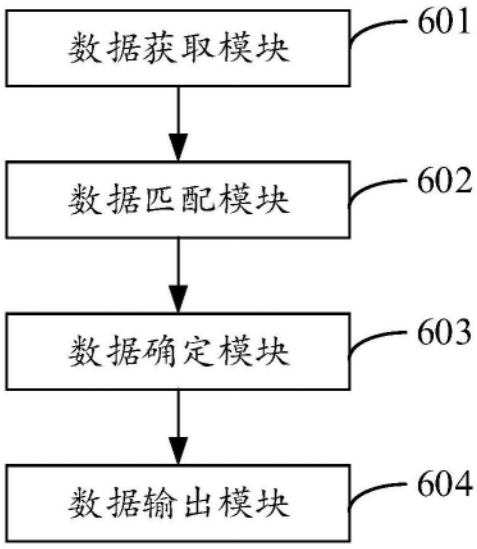

数据获取模块
601
数据匹配模块
602
数据确定模块
603
数据输出模块
604

图6

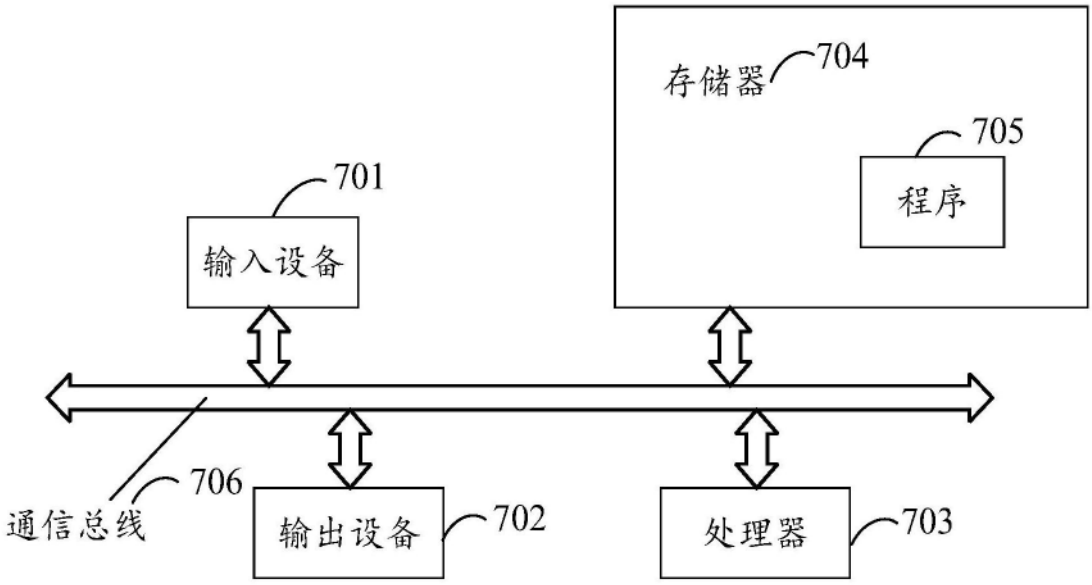

存储器
704
705
程序
701
输入设备
706
通信总线
输出设备
702
处理器
703

图7