



(12) 发明专利申请

(10) 申请公布号 CN 103914533 A

(43) 申请公布日 2014. 07. 09

(21) 申请号 201410127121. 2

(22) 申请日 2014. 03. 31

(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 陈志杰 江琳 韩帅 刘晓春
何仁清 李小松 郭伟 罗斌

(74) 专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201
代理人 宋合成

(51) Int. Cl.
G06F 17/30(2006. 01)

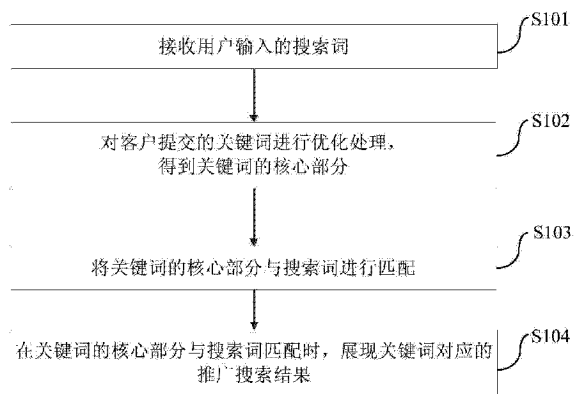
权利要求书3页 说明书11页 附图3页

(54) 发明名称

推广搜索结果的展现方法和装置

(57) 摘要

本发明提出一种推广搜索结果的提供方法和装置,为提高推广搜索结果页的利用效率。该推广搜索结果页的提供方法包括:接收用户输入的搜索词;对客户提交的关键词进行优化处理,得到关键词的核心部分;将关键词的核心部分与搜索词进行匹配;在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。该方法能够提高推广搜索结果页的利用效率,通过推广搜索结果页提供更多的信息。



1. 一种推广搜索结果的展现方法,其特征在于,包括:
接收用户输入的搜索词;
对客户提交的关键词进行优化处理,得到所述关键词的核心部分;
将所述关键词的核心部分与所述搜索词进行匹配;
在所述关键词的核心部分与所述搜索词匹配时,展现所述关键词对应的推广搜索结果。
2. 根据权利要求1所述的方法,其特征在于,所述对客户提交的关键词进行优化处理,得到所述关键词的核心部分,具体包括:
对所述关键词进行缩减处理,获取所述关键词对应的缩减词;
计算所述缩减词与所述关键词的语义相关性数值;以及
当所述语义相关性数值大于预设的第一阈值时,确定所述缩减词为所述关键词的核心部分。
3. 根据权利要求1所述的方法,其特征在于,所述将所述关键词的核心部分与所述搜索词进行匹配,具体包括:
对所述搜索词进行预处理,获取预处理后的搜索词;
获取所述预处理后的搜索词与所述关键词的核心部分的业务相关性数值;以及
当所述业务相关性数值大于预设的第二阈值时,确定所述关键词的核心部分与所述搜索词匹配。
4. 根据权利要求2所述的方法,其特征在于,所述计算所述缩减词与所述关键词的语义相关性数值,具体包括:
获取第一向量和第二向量,所述第一向量是所述关键词对应的词向量,所述第二向量是所述缩减词对应的词向量;
计算所述第一向量和第二向量的相似度数值,将所述相似度数值确定为所述语义相关性数值。
5. 根据权利要求3所述的方法,其特征在于,所述获取所述搜索词与所述关键词的核心部分的业务相关性数值,具体包括:
根据预先得到的业务相关性模型,获取所述搜索词与所述关键词的核心部分的业务相关性数值,所述业务相关性模型为预先根据数据样本的相关度进行训练后得到的。
6. 根据权利要求2所述的方法,其特征在于,所述对所述关键词进行缩减处理,获取所述关键词对应的缩减词,具体包括:
对所述关键词进行预处理,以获取所述关键词对应的多个分词;
对所述多个分词进行去冗余处理,以获取所述缩减词。
7. 根据权利要求6所述的方法,其特征在于,所述对所述多个分词进行去冗余处理,具体包括:
根据所述关键词的语义,去除所述多个分词中与上下文无关的分词;以及
分析所述多个分词的语义,去除所述多个分词中语义重复的分词。
8. 根据权利要求6所述的方法,其特征在于,所述预处理包括如下项中的至少一项:
切分词、专有名词识别。
9. 根据权利要求3所述的方法,其特征在于,所述对所述搜索词进行预处理,包括:

对所述搜索词进行切词处理,和 / 或,纠错处理。

10. 根据权利要求 1 所述的方法,其特征在于,所述展现所述关键词对应的推广搜索结果,具体包括:

根据预设的规则对所述搜索结果进行排序,得到筛选后的推广搜索结果;
展现所述筛选后的推广搜索结果。

11. 一种推广搜索结果的展现装置,其特征在于,包括:

接收模块,用于接收用户输入的搜索词;

优化模块,用于对客户提交的关键词进行优化处理,得到所述关键词的核心部分;

匹配模块,用于将所述关键词的核心部分与所述搜索词进行匹配;

展现模块,用于在所述关键词的核心部分与所述搜索词匹配时,展现所述关键词对应的推广搜索结果。

12. 根据权利要求 11 所述的装置,其特征在于,所述优化模块具体包括:

缩减子模块,用于对所述关键词进行缩减处理,以获取所述关键词对应的缩减词;

计算子模块,用于计算所述缩减词与所述关键词的语义相关性数值;以及

第一确定子模块,用于在所述语义相关性数值大于预设的第一阈值时,确定所述缩减词为所述关键词的核心部分。

13. 根据权利要求 11 所述的装置,其特征在于,所述匹配模块具体包括:

预处理子模块,用于对所述搜索词进行预处理,以获取预处理后的搜索词;

获取子模块,用于获取所述预处理后的搜索词与所述关键词的核心部分的业务相关性数值;以及

第二确定子模块,用于当所述业务相关性数值大于预设的第二阈值时,确定所述关键词的核心部分与所述搜索词匹配。

14. 根据权利要求 12 所述的装置,其特征在于,所述计算子模块具体包括:

获取单元,用于获取第一向量和第二向量,所述第一向量是所述关键词对应的词向量,所述第二向量是所述缩减词对应的词向量;

计算单元,用于计算所述第一向量和第二向量的相似度数值,将所述相似度数值确定为所述语义相关性数值。

15. 根据权利要求 13 所述的装置,其特征在于,所述获取子模块具体用于:

根据预先得到的业务相关性模型,获取所述搜索词语所述关键词的核心部分的业务相关性数值,所述业务相关性模型为预先根据数据样本的相关度进行训练后得到的。

16. 根据权利要求 12 所述的装置,其特征在于,所述缩减子模块具体包括:

预处理单元,用于对所述关键词进行预处理,以获取所述关键词对应的多个分词;

去冗余单元,用于对所述多个分词进行去冗余处理,以获取所述缩减词。

17. 根据权利要求 16 所述的装置,其特征在于,所述去冗余单元具体用于:

根据所述关键词的语义,去除所述多个分词中与上下文无关的分词;以及

分析所述多个分词的语义,去除所述多个分词中语义重复的分词。

18. 根据权利要求 16 所述的装置,其特征在于,所述预处理单元具体用于对所述关键词进行切分词、专有名词识别中的至少一项处理。

19. 根据权利要求 13 所述的装置,其特征在于,所述预处理子模块具体用于对所述搜

索词进行切词处理,和 / 或,纠错处理。

20. 根据权利要求 11 所述的装置,其特征在于,所述展现模块具体包括:

排序子模块,用于根据预设的规则对所述推广搜索结果进行排序,得到筛选后的推广搜索结果;

展现子模块,用于展现所述筛选后的推广搜索结果。

推广搜索结果的展现方法和装置

技术领域

[0001] 本发明涉及搜索技术领域,尤其涉及一种推广搜索结果的展现方法和装置。

背景技术

[0002] 目前的推广搜索引擎主流技术是关键词短语匹配技术,其主要根据搜索词与关键词是否字面匹配或同义匹配,即搜索词字面包含关键词或关键词的同义词。但是,采用这种短语匹配技术,推广者购买的很长关键词很难被匹配到,导致其广告无法被展现。由于推广客户不知道如何对关键词进行优化,导致其为增加展现量,只能不断提交新词,大大增加了其时间和人力成本。

发明内容

[0003] 本发明旨在至少在一定程度上解决相关技术中的技术问题之一。

[0004] 为此,本发明的一个目的在于提出一种推广搜索结果的展现方法,该方法可以更充分利用推广搜索结果页,不需要网页跳转就可以在推广搜索结果页中获取更多信息,提高资源利用效率。

[0005] 本发明的另一个目的在于提出一种推广搜索结果的展现装置。

[0006] 为达到上述目的,本发明第一方面实施例提出的推广搜索结果的提供方法,包括:接收用户输入的搜索词;对客户提交的关键词进行优化处理,得到关键词的核心部分;将关键词的核心部分与搜索词进行匹配;在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。

[0007] 本发明第一方面实施例提出的推广搜索结果的提供方法,通过对客户提交的关键词进行优化处理,得到关键词的核心部分,并将关键词的核心部分与搜索词进行匹配,避免了关键词过长导致的匹配率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0008] 为达到上述目的,本发明第二方面实施例提出的推广搜索结果的展现装置,包括:接收模块,用于接收用户输入的搜索词;优化模块,用于对客户提交的关键词进行优化处理,得到关键词的核心部分;匹配模块,用于将关键词的核心部分与搜索词进行匹配;展现模块,用于在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。

[0009] 本发明第二方面实施例提出的推广搜索结果的展现装置,通过对客户提交的关键词进行优化处理,得到关键词的核心部分,并将关键词的核心部分与搜索词进行匹配,避免了关键词过长导致的匹配率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0010] 为达到上述目的,本发明第三方面实施例提出的客户端设备,包括:壳体、处理器、存储器、电路板和电源电路,其中,电路板安置在壳体围成的空间内部,处理器和存储器设

置在电路板上；电源电路，用于为客户端设备的各个电路或器件供电；存储器用于存储可执行程序代码；处理器通过读取存储器中存储的可执行程序代码来运行与可执行程序代码对应的程序，以用于执行以下步骤：接收用户输入的搜索词；对客户提交的关键词进行优化处理，得到关键词的核心部分；将关键词的核心部分与搜索词进行匹配；在关键词的核心部分与搜索词匹配时，展现关键词对应的搜索结果。

[0011] 本发明第三方面实施例提出的客户端设备，通过对客户提交的关键词进行优化处理，得到关键词的核心部分，并将关键词的核心部分与搜索词进行匹配，避免了关键词过长导致的匹配率低，充分利用搜索结果页，不需要频繁更换搜索词就可以在搜索结果页中获取更多信息，优化了搜索结果的展现，提高了资源利用效率，提升了用户体验。本发明附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本发明的实践了解到。

附图说明

[0012] 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

[0013] 图 1 是本发明一个实施例提出的一种推广搜索结果的提供方法的流意图；

[0014] 图 2 是本发明又一个实施例提出的一种推广搜索结果的提供方法的流程图；

[0015] 图 3 是本发明一个实施例提出的一种推广搜索结果的提供装置的结构框图；以及

[0016] 图 4 是本发明又一个实施例提出的一种推广搜索结果的提供装置的结构框图。

具体实施方式

[0017] 下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。相反，本发明的实施例包括落入所附加权利要求书的精神和内涵范围内的所有变化、修改和等同物。

[0018] 下面参考附图描述根据本发明实施例的推广搜索结果的展现方法和装置。

[0019] 图 1 为本发明一个实施例的推广搜索结果的展现方法的流程图，该方法包括：

[0020] S101：接收用户输入的搜索词。

[0021] 具体地，用户可通过在电脑、移动电子设备等终端设备的搜索引擎或其他具备搜索功能的软件中输入感兴趣的搜索词，以获取想要搜索的内容。

[0022] S102：对客户提交的关键词进行优化处理，得到关键词的核心部分。

[0023] 举例而言，以搜索推广为例，客户即为搜索推广内容的提供者，客户可将希望推广的内容（如品牌、广告、网站链接等）及其关键词提交给搜索引擎，当用户搜索相应的关键词时，搜索引擎即可将客户想要推广的内容展现给搜索用户，以达到推广的目的。

[0024] 具体地，在本发明的一个实施例中，搜索引擎可对客户提交的关键词进行优化处理，例如，对字数过多的长关键词进行拆分或缩减，去除关键词中的冗余部分等，从而得到关键词中与客户所要展现或推广的内容相关性更高的核心部分。其中，核心部分可以由一个或多个词或短语构成。

[0025] S103 :将关键词的核心部分与搜索词进行匹配。

[0026] 具体地,可根据接收到的搜索词查找关键词的核心部分,并根据关键词的核心部分与搜索词的相关性确定关键词的核心部分是否与搜索词匹配。

[0027] 在本发明的一个实施例中,可预先对用户输入的搜索词进行切词和 / 或纠错等预处理,然后再将预处理后的搜索词与关键词的核心部分进行匹配。

[0028] S104 :在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。

[0029] 具体地,当查询到与当前搜索词匹配的关键词的核心部分时,搜索引擎可根据预设的规则(例如关键词的核心部分与搜索词的匹配程度等)对推广搜索结果进行排序和筛选,然后将对应的排序后的推广搜索结果展现给用户。

[0030] 本发明实施例的推广搜索结果的展现方法,通过对客户提交的关键词进行优化处理,得到关键词的核心部分,并将关键词的核心部分与搜索词进行匹配,避免了关键词过长导致的匹配率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0031] 图 2 为根据本发明又一个实施例的推广搜索结果的展现方法的流程图。为了进一步提升用户体验,可先对关键词和搜索词分别进行处理,然后将处理得到的关键词的核心部分与预处理后的搜索词进行匹配,以提高搜索匹配率。如图 2 所示,该推广搜索结果的展现方法包括:

[0032] S201 :接收用户输入的搜索词。

[0033] 具体地,用户可通过在电脑、移动电子设备等终端设备的搜索引擎或其他具备搜索功能的软件中输入感兴趣的搜索词,以获取想要搜索的内容。

[0034] S202 :对客户提交的关键词进行缩减处理,获取关键词对应的缩减词。

[0035] 举例而言,以搜索推广为例,客户即为搜索推广内容的提供者,客户可将希望推广的内容(如品牌、广告、网站链接等)及其关键词提交给搜索引擎,当用户搜索相应的关键词时,搜索引擎即可将客户想要推广的内容展现给搜索用户,以达到推广的目的。

[0036] 具体地,可对客户提交的关键词进行预处理,以获取关键词对应的多个分词。其中,预处理包括切分词、专有名词识别中的至少一种处理。例如,将搜索词“如家快捷酒店北京店的节假日收费标准”切分为“如家快捷酒店”、“北京店的”、“节假日”、“收费标准”或更多项,并识别其中的专有名词“如家快捷酒店”和“北京”。

[0037] 另外,还可以对多个分词进行去冗余处理,以获取缩减词。其中,对多个分词进行去冗余处理具体包括根据关键词的语义,去除多个分词中与上下文无关的分词,例如可通过自然语言数据处理等方法对多个分词的成分进行分析,去除多个分词中的停用词(包括无实际意义的功能词,如“的”、“了”等,以及对搜索贡献度较低的词汇词,如“想要”、“希望”等);还可以分析多个分词的语义,去除多个分词中语义重复的分词。具体地,以“如家快捷酒店”为例,可通过用户行为数据等资料挖掘可省略的词对,结合上下文,判断“快捷”为冗余成分,并进行去除,得到缩减词“如家酒店”。

[0038] S203 :计算缩减词与关键词的语义相关性数值。

[0039] 在本发明的一个实施例中,可获取第一向量和第二向量,其中,第一向量是关键词对应的词向量,第二向量是缩减词对应的词向量,并计算第一向量和第二向量的相似度数

值,然后将计算得到的相似度数值确定为缩减词与关键词的语义相关性数值。具体地,举例而言,可利用词向量模型,计算出关键词对应的词向量 $\text{Vector}_W(x_1, x_2, \dots, x_n)$,和缩减词对应的词向量 $\text{Vector}_W\text{-reduction}(y_1, y_2, \dots, y_n)$,其中 n 为词向量的维数,并计算两个词向量的相似度数值,计算公式如下:

[0040] $\text{Sim}_{w\text{-reduction}} =$

[0041] $(x_1y_1+x_2y_2+\dots+x_ny_n) / [(x_1^2+x_2^2+\dots+x_n^2)(y_1^2+y_2^2+\dots+y_n^2)]^{1/2}$

[0042] 即可将得到的相似度数值作为缩减词与关键词的语义相关性数值。

[0043] S204:当语义相关性数值大于预设的第一阈值时,确定缩减词为关键词的核心部分。

[0044] 其中,预设的第一阈值可由搜索引擎设定。可将计算得到的缩减词与关键词的语义相似度数值与预设的第一阈值进行比较,当语义相关性数值大于预设的第一阈值时,确定对应的缩减词为关键词的核心部分。其中,关键词的核心部分可以为一个或多个缩减词。

[0045] S205:对搜索词进行预处理,获取预处理后的搜索词。

[0046] 其中,预处理包括对搜索词进行切词处理和/或纠错处理。

[0047] 具体地,当用户输入的搜索词过长时,很可能会造成无法匹配到对应的关键词或关键词的核心部分,此时可通过对搜索词进行切词处理,将搜索词拆分成几个较短的搜索词,而不改变搜索词的原意,从而提高搜索匹配率。

[0048] 另外,由于用户在输入搜索词时,有可能由于输入错误或信息模糊导致输入的搜索词与常见的搜索词不同,导致难以匹配到相关性较高的推广搜索结果,因此可对搜索词进行纠错处理以提高搜索效率,例如,当用户输入“儒家酒店”时,可提示用户想要搜索的是否为同音词“如家酒店”,并提供仍然搜索“儒家酒店”的指引选项。

[0049] 需要理解的是,具体的切分方法和纠错方法有多种,在此不再一一列举。

[0050] S206:获取预处理后的搜索词与关键词的核心部分的业务相关性数值。

[0051] 具体地,在本发明的一个实施例中,可根据预先得到的业务相关性模型,提取搜索词与关键词的核心部分的行业属性、意图属性等特征,从而计算搜索词与关键词的核心部分的业务相关性数值。

[0052] 其中,业务相关性模型为预先根据大量数据样本的相关度进行训练后得到的(例如利用机器学习方法等),训练时采用的数据样本是用 1 分表示业务相关,-1 分表示业务不相关。

[0053] 应当理解的是,业务相关性模型的具体实现可以采用相关技术实现,本发明实施例不再赘述。

[0054] S207:当业务相关性数值大于预设的第二阈值时,确定关键词的核心部分与搜索词匹配。

[0055] 具体地,可以将计算得到的搜索词与关键词的核心部分的业务相关性数值与预设的第二阈值进行比较,当业务相关性数值大于预设的第二阈值时,确定关键词的核心部分与搜索词匹配。其中,预设的第二阈值可由搜索引擎设定。

[0056] S208:在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。

[0057] 具体地,当查询到与当前搜索词匹配的关键词的核心部分时,可根据预设的规则(例如关键词的核心部分与搜索词的匹配程度等)对推广搜索结果进行排序,得到筛选后的

推广搜索结果,然后将对应的排序后的推广搜索结果展现给用户。

[0058] 本发明实施例的推广搜索结果的展现方法,通过对客户提交的关键词进行缩减处理,并通过计算缩减词与关键词的语义相关性得到关键词的核心部分,然后根据业务相关性将关键词的核心部分与预处理后的搜索词进行匹配,使关键词更容易被搜索到,提高了搜索匹配的准确率,同时,对搜索词进行预处理也避免了搜索词过长导致的匹配准确率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多、更准确的信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0059] 为了实现上述实施例,本发明还提出一种推广搜索结果的展现装置。

[0060] 图3为根据本发明一个实施例的推广搜索结果的展现装置的结构框图。如图3所示,该推广搜索结果的展现装置包括:接收模块100、优化模块200、匹配模块300和展现模块400。

[0061] 具体地,接收模块100用于接收用户输入的搜索词。更具体地,用户可通过电脑、移动电子设备等终端设备的搜索引擎或其他具备搜索功能的软件将感兴趣的搜索词输入到接收模块100,以获取想要搜索的内容。

[0062] 优化模块200用于对客户提交的关键词进行优化处理,得到关键词的核心部分。举例而言,以搜索推广为例,客户即为搜索推广内容的提供者,客户可将希望推广的内容(如品牌、广告、网站链接等)及其关键词提交给搜索引擎,当用户搜索相应的关键词时,搜索引擎即可将客户想要推广的内容展现给搜索用户,以达到推广的目的。更具体地,在本发明的一个实施例中,优化模块200可对接收模块100接收到的客户提交的关键词进行优化处理,例如,对字数过多的长关键词进行拆分或缩减,去除关键词中的冗余部分等,从而得到关键词中与客户所要展现或推广的内容相关性更高的核心部分。其中,核心部分可以由一个或多个词或短语构成。

[0063] 匹配模块300用于将关键词的核心部分与搜索词进行匹配。更具体地,匹配模块300可根据接收到的搜索词查找关键词的核心部分,并根据关键词的核心部分与搜索词的相关性确定关键词的核心部分是否与搜索词匹配。在本发明的一个实施例中,匹配模块300可预先对用户输入的搜索词进行切词和/或纠错等预处理,然后再将预处理后的搜索词与关键词的核心部分进行匹配。

[0064] 展现模块400用于在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。更具体地,当匹配模块300查询到与当前搜索词匹配的关键词的核心部分时,展现模块400可根据预设的规则(例如关键词的核心部分与搜索词的匹配程度等)对推广搜索结果进行排序和筛选,然后将对应的排序后的推广搜索结果展现给用户。

[0065] 本发明实施例的推广搜索结果的展现装置,通过对客户提交的关键词进行优化处理,得到关键词的核心部分,并将关键词的核心部分与搜索词进行匹配,避免了关键词过长导致的匹配率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0066] 图4为根据本发明另一个实施例的推广搜索结果的展现装置的结构框图。如图4所示,该推广搜索结果的展现装置包括:接收模块100、优化模块200、缩减子模块210、预处理单元211、去冗余单元212、计算子模块220、获取单元221、计算单元222、第一确定子模块

230、匹配模块 300、预处理子模块 310、获取子模块 320、第二确定子模块 330、展现模块 400、排序子模块 410 和展现子模块 420。其中,优化模块 200 包括缩减子模块 210、计算子模块 220 和第一确定子模块 230,缩减子模块 210 包括预处理单元 211 和去冗余单元 212,计算子模块 220 包括获取单元 221 和计算单元 222,匹配模块 300 包括预处理子模块 310、获取子模块 320 和第二确定子模块 330,展现模块 400 包括排序子模块 410 和展现子模块 420。

[0067] 具体地,缩减子模块 210 用于对关键词进行缩减处理,以获取关键词对应的缩减词。更具体地,缩减子模块 210 可通过预处理单元 211 可对客户提交的关键词进行预处理,以获取关键词对应的多个分词。其中,预处理包括切分词、专有名词识别中的至少一种处理。例如,将搜索词“如家快捷酒店北京店的节假日收费标准”切分为“如家快捷酒店”、“北京店的”、“节假日”、“收费标准”或更多项,并识别其中的专有名词“如家快捷酒店”和“北京”。

[0068] 另外,缩减子模块 210 还可以通过去冗余单元 212 对多个分词进行去冗余处理,以获取缩减词。其中,对多个分词进行去冗余处理具体包括根据关键词的语义,去除多个分词中与上下文无关的分词,例如可通过自然语言数据处理等方法对多个分词的成分进行分析,去除多个分词中的停用词(包括无实际意义的功能词,如“的”、“了”等,以及对搜索贡献度较低的词汇词,如“想要”、“希望”等);还可以分析多个分词的语义,去除多个分词中语义重复的分词。更具体地,以“如家快捷酒店”为例,可通过用户行为数据等资料挖掘可省略的词对,结合上下文,判断“快捷”为冗余成分,并进行去除,得到缩减词“如家酒店”。

[0069] 计算子模块 220 用于计算缩减词与关键词的语义相关性数值。更具体地,计算子模块 220 可通过获取单元 221 获取第一向量和第二向量,其中,第一向量是关键词对应的词向量,第二向量是缩减词对应的词向量,并通过计算单元 222 计算第一向量和第二向量的相似度数,然后将计算得到的相似度数确定为缩减词与关键词的语义相关性数值。举例而言,在本发明的一个实施例中,计算子模块 220 可利用词向量模型,计算出关键词对应的词向量 $\text{Vector_W}(x_1, x_2, \dots, x_n)$, 和缩减词对应的词向量 $\text{Vector_W_reduction}(y_1, y_2, \dots, y_n)$, 其中 n 为词向量的维数,并计算两个词向量的相似度数,计算公式如下:

[0070] $\text{Sim}_{w_reduction} =$

[0071] $(x_1y_1+x_2y_2+\dots+x_ny_n) / [(x_1^2+x_2^2+\dots+x_n^2)(y_1^2+y_2^2+\dots+y_n^2)]^{1/2}$

[0072] 即可将得到的相似度数作为缩减词与关键词的语义相关性数值。

[0073] 第一确定子模块 230 用于在语义相关性数值大于预设的第一阈值时,确定缩减词为关键词的核心部分。其中,预设的第一阈值可由搜索引擎设定。第一确定子模块 230 可将计算得到的缩减词与关键词的语义相似度数与预设的第一阈值进行比较,当语义相关性数值大于预设的第一阈值时,确定对应的缩减词为关键词的核心部分。其中,关键词的核心部分可以为一个或多个缩减词。

[0074] 预处理子模块 310 用于对搜索词进行预处理,以获取预处理后的搜索词。其中,预处理包括对搜索词进行切词处理和 / 或纠错处理。更具体地,当用户输入的搜索词过长时,很可能会造成无法匹配到对应的关键词或关键词的核心部分,此时可通过预处理子模块 310 对搜索词进行切词处理,将搜索词拆分成几个较短的搜索词,而不改变搜索词的原意,从而提高搜索匹配率。另外,由于用户在输入搜索词时,有可能由于输入错误或信息模糊导致输入的搜索词与常见的搜索词不同,导致难以匹配到相关性较高的推广搜索结果,

因此预处理子模块 310 还可对搜索词进行纠错处理以提高搜索效率,例如,当用户输入“儒家酒店”时,可提示用户想要搜索的是否为同音词“如家酒店”,并提供仍然搜索“儒家酒店”的指引选项。

[0075] 需要理解的是,具体的切分方法和纠错方法有多种,在此不再一一列举。

[0076] 获取子模块 320 用于获取预处理后的搜索词与关键词的核心部分的业务相关性数值。在本发明的一个实施例中,获取子模块 320 可根据预先得到的业务相关性模型,提取搜索词与关键词的核心部分的行业属性、意图属性等特征,从而计算搜索词与关键词的核心部分的业务相关性数值。其中,业务相关性模型为预先根据大量数据样本的相关度进行训练后得到的(例如利用机器学习方法等),训练时采用的数据样本是用 1 分表示业务相关,-1 分表示业务不相关。

[0077] 应当理解的是,业务相关性模型的具体实现可以采用相关技术实现,本发明实施例不再赘述。

[0078] 第二确定子模块 330 用于当业务相关性数值大于预设的第二阈值时,确定关键词的核心部分与搜索词匹配。更具体地,第二确定子模块 330 可以将计算得到的搜索词与关键词的核心部分的业务相关性数值与预设的第二阈值进行比较,当业务相关性数值大于预设的第二阈值时,确定关键词的核心部分与搜索词匹配。其中,预设的第二阈值可由搜索引擎设定。

[0079] 排序子模块 410 用于根据预设的规则对推广搜索结果进行排序,得到筛选后的推广搜索结果。

[0080] 展现子模块 420 用于展现筛选后的推广搜索结果。

[0081] 更具体地,当查询到与当前搜索词匹配的关键词的核心部分时,排序子模块 410 可根据预设的规则(例如关键词的核心部分与搜索词的匹配程度等)对推广搜索结果进行排序,得到筛选后的推广搜索结果,然后通过展现子模块 420 将对应的排序后的推广搜索结果展现给用户。

[0082] 应当理解的是,在本发明的一个实施例中,上述任一实施例的推广搜索结果的展现装置均可以为具有上述功能模块的终端设备,如电脑主机等,具体在此不再一一列举。

[0083] 本发明实施例还提供了一种推广搜索结果的展现装置,能够对客户提交的关键词进行缩减处理,通过计算缩减词与关键词的语义相关性得到关键词的核心部分,并根据业务相关性将关键词的核心部分与预处理后的搜索词进行匹配,使关键词更容易被搜索到,提高了搜索匹配的准确率,同时,对搜索词进行预处理也避免了搜索词过长导致的匹配准确率低,充分利用了推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多、更准确的信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0084] 为达到上述目的,本发明实施例提出的客户端设备,包括:壳体、处理器、存储器、电路板和电源电路,其中,电路板安置在壳体围成的空间内部,处理器和存储器设置在电路板上;电源电路,用于为客户端设备的各个电路或器件供电;存储器用于存储可执行程序代码;处理器通过读取存储器中存储的可执行程序代码来运行与可执行程序代码对应的程序,以用于执行以下步骤:

[0085] S101':接收用户输入的搜索词。

[0086] 具体地,用户可通过在电脑、移动电子设备等终端设备的搜索引擎或其他具备搜索功能的软件中输入感兴趣的搜索词,以获取想要搜索的内容。

[0087] S102':对客户提交的关键词进行优化处理,得到关键词的核心部分。

[0088] 举例而言,以搜索推广为例,客户即为搜索推广内容的提供者,客户可将希望推广的内容(如品牌、广告、网站链接等)及其关键词提交给搜索引擎,当用户搜索相应的关键词时,搜索引擎即可将客户想要推广的内容展现给搜索用户,以达到推广的目的。

[0089] 具体地,在本发明的一个实施例中,搜索引擎可对客户提交的关键词进行优化处理,例如,对字数过多的长关键词进行拆分或缩减,去除关键词中的冗余部分等,从而得到关键词中与客户所要展现或推广的内容相关性更高的核心部分。其中,核心部分可以由一个或多个词或短语构成。

[0090] S103':将关键词的核心部分与搜索词进行匹配。

[0091] 具体地,可根据接收到的搜索词查找关键词的核心部分,并根据关键词的核心部分与搜索词的相关性确定关键词的核心部分是否与搜索词匹配。

[0092] 在本发明的一个实施例中,可预先对用户输入的搜索词进行切词和/或纠错等预处理,然后再将预处理后的搜索词与关键词的核心部分进行匹配。

[0093] S104':在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。

[0094] 具体地,当查询到与当前搜索词匹配的关键词的核心部分时,搜索引擎可根据预设的规则(例如关键词的核心部分与搜索词的匹配程度等)对推广搜索结果进行排序和筛选,然后将对应的排序后的推广搜索结果展现给用户。

[0095] 本发明实施例的推广搜索结果的展现方法,通过对客户提交的关键词进行优化处理,得到关键词的核心部分,并将关键词的核心部分与搜索词进行匹配,避免了关键词过长导致的匹配率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0096] 另一实施例中,处理器通过读取存储器中存储的可执行程序代码来运行与可执行程序代码对应的程序,以用于执行以下步骤:

[0097] S201':接收用户输入的搜索词。

[0098] 具体地,用户可通过在电脑、移动电子设备等终端设备的搜索引擎或其他具备搜索功能的软件中输入感兴趣的搜索词,以获取想要搜索的内容。

[0099] S202':对客户提交的关键词进行缩减处理,获取关键词对应的缩减词。

[0100] 举例而言,以搜索推广为例,客户即为搜索推广内容的提供者,客户可将希望推广的内容(如品牌、广告、网站链接等)及其关键词提交给搜索引擎,当用户搜索相应的关键词时,搜索引擎即可将客户想要推广的内容展现给搜索用户,以达到推广的目的。

[0101] 具体地,可对客户提交的关键词进行预处理,以获取关键词对应的多个分词。其中,预处理包括切分词、专有名词识别中的至少一种处理。例如,将搜索词“如家快捷酒店北京店的节假日收费标准”切分为“如家快捷酒店”、“北京店的”、“节假日”、“收费标准”或更多项,并识别其中的专有名词“如家快捷酒店”和“北京”。

[0102] 另外,还可以对多个分词进行去冗余处理,以获取缩减词。其中,对多个分词进行去冗余处理具体包括根据关键词的语义,去除多个分词中与上下文无关的分词,例如可通

过自然语言数据处理等方法对多个分词的成分进行分析,去除多个分词中的停用词(包括无实际意义的功能词,如“的”、“了”等,以及对搜索贡献度较低的词汇词,如“想要”、“希望”等);还可以分析多个分词的语义,去除多个分词中语义重复的分词。具体地,以“如家快捷酒店”为例,可通过用户行为数据等资料挖掘可省略的词对,结合上下文,判断“快捷”为冗余成分,并进行去除,得到缩减词“如家酒店”。

[0103] S203':计算缩减词与关键词的语义相关性数值。

[0104] 在本发明的一个实施例中,可获取第一向量和第二向量,其中,第一向量是关键词对应的词向量,第二向量是缩减词对应的词向量,并计算第一向量和第二向量的相似度数值,然后将计算得到的相似度数值确定为缩减词与关键词的语义相关性数值。具体地,举例而言,可利用词向量模型,计算出关键词对应的词向量 $\text{Vector_W}(x_1, x_2, \dots, x_n)$,和缩减词对应的词向量 $\text{Vector_W_reduction}(y_1, y_2, \dots, y_n)$,其中 n 为词向量的维数,并计算两个词向量的相似度数值,计算公式如下:

[0105] $\text{Sim}_{w_reduction} =$

[0106] $(x_1y_1+x_2y_2+\dots+x_ny_n) / [(x_1^2+x_2^2+\dots+x_n^2)(y_1^2+y_2^2+\dots+y_n^2)]^{1/2}$

[0107] 即可将得到的相似度数值作为缩减词与关键词的语义相关性数值。

[0108] S204':当语义相关性数值大于预设的第一阈值时,确定缩减词为关键词的核心部分。

[0109] 其中,预设的第一阈值可由搜索引擎设定。可将计算得到的缩减词与关键词的语义相似度数值与预设的第一阈值进行比较,当语义相关性数值大于预设的第一阈值时,确定对应的缩减词为关键词的核心部分。其中,关键词的核心部分可以为一个或多个缩减词。

[0110] S205':对搜索词进行预处理,获取预处理后的搜索词。

[0111] 其中,预处理包括对搜索词进行切词处理和/或纠错处理。

[0112] 具体地,当用户输入的搜索词过长时,很可能会造成无法匹配到对应的关键词或关键词的核心部分,此时可通过对搜索词进行切词处理,将搜索词拆分成几个较短的搜索词,而不改变搜索词的原意,从而提高搜索匹配率。

[0113] 另外,由于用户在输入搜索词时,有可能由于输入错误或信息模糊导致输入的搜索词与常见的搜索词不同,导致难以匹配到相关性较高的推广搜索结果,因此可对搜索词进行纠错处理以提高搜索效率,例如,当用户输入“儒家酒店”时,可提示用户想要搜索的是否为同音词“如家酒店”,并提供仍然搜索“儒家酒店”的指引选项。

[0114] 需要理解的是,具体的切分方法和纠错方法有多种,在此不再一一列举。

[0115] S206':获取预处理后的搜索词与关键词的核心部分的业务相关性数值。

[0116] 具体地,在本发明的一个实施例中,可根据预先得到的业务相关性模型,提取搜索词与关键词的核心部分的行业属性、意图属性等特征,从而计算搜索词与关键词的核心部分的业务相关性数值。

[0117] 其中,业务相关性模型为预先根据大量数据样本的相关度进行训练后得到的(例如利用机器学习方法等),训练时采用的数据样本是用 1 分表示业务相关,-1 分表示业务不相关。

[0118] 应当理解的是,业务相关性模型的具体实现可以采用相关技术实现,本发明实施例不再赘述。

[0119] S207' :当业务相关性数值大于预设的第二阈值时,确定关键词的核心部分与搜索词匹配。

[0120] 具体地,可以将计算得到的搜索词与关键词的核心部分的业务相关性数值与预设的第二阈值进行比较,当业务相关性数值大于预设的第二阈值时,确定关键词的核心部分与搜索词匹配。其中,预设的第二阈值可由搜索引擎设定。

[0121] S208' :在关键词的核心部分与搜索词匹配时,展现关键词对应的推广搜索结果。

[0122] 具体地,当查询到与当前搜索词匹配的关键词的核心部分时,可根据预设的规则(例如关键词的核心部分与搜索词的匹配程度等)对推广搜索结果进行排序,得到筛选后的推广搜索结果,然后将对应的排序后的推广搜索结果展现给用户。

[0123] 本发明实施例的推广搜索结果的展现方法,通过对客户提交的关键词进行缩减处理,并通过计算缩减词与关键词的语义相关性得到关键词的核心部分,然后根据业务相关性将关键词的核心部分与预处理后的搜索词进行匹配,使关键词更容易被搜索到,提高了搜索匹配的准确率,同时,对搜索词进行预处理也避免了搜索词过长导致的匹配准确率低,充分利用推广搜索结果页,不需要频繁更换搜索词就可以在推广搜索结果页中获取更多、更准确的信息,优化了推广搜索结果的展现,提高了资源利用效率,提升了用户体验。

[0124] 需要说明的是,在本发明的描述中,术语“第一”、“第二”等仅用于描述目的,而不能理解为指示或暗示相对重要性。此外,在本发明的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0125] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本发明的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本发明的实施例所属技术领域的技术人员所理解。

[0126] 应当理解,本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如,如果用硬件来实现,和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或他们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0127] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0128] 此外,在本发明各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读取存储介质中。

[0129] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0130] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示

例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

[0131] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

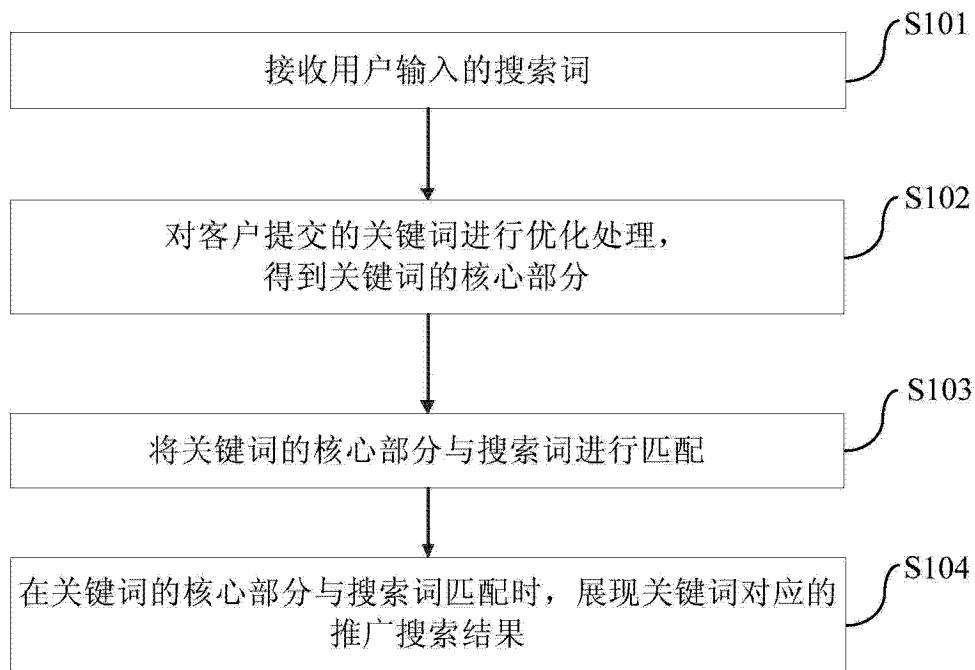


图 1

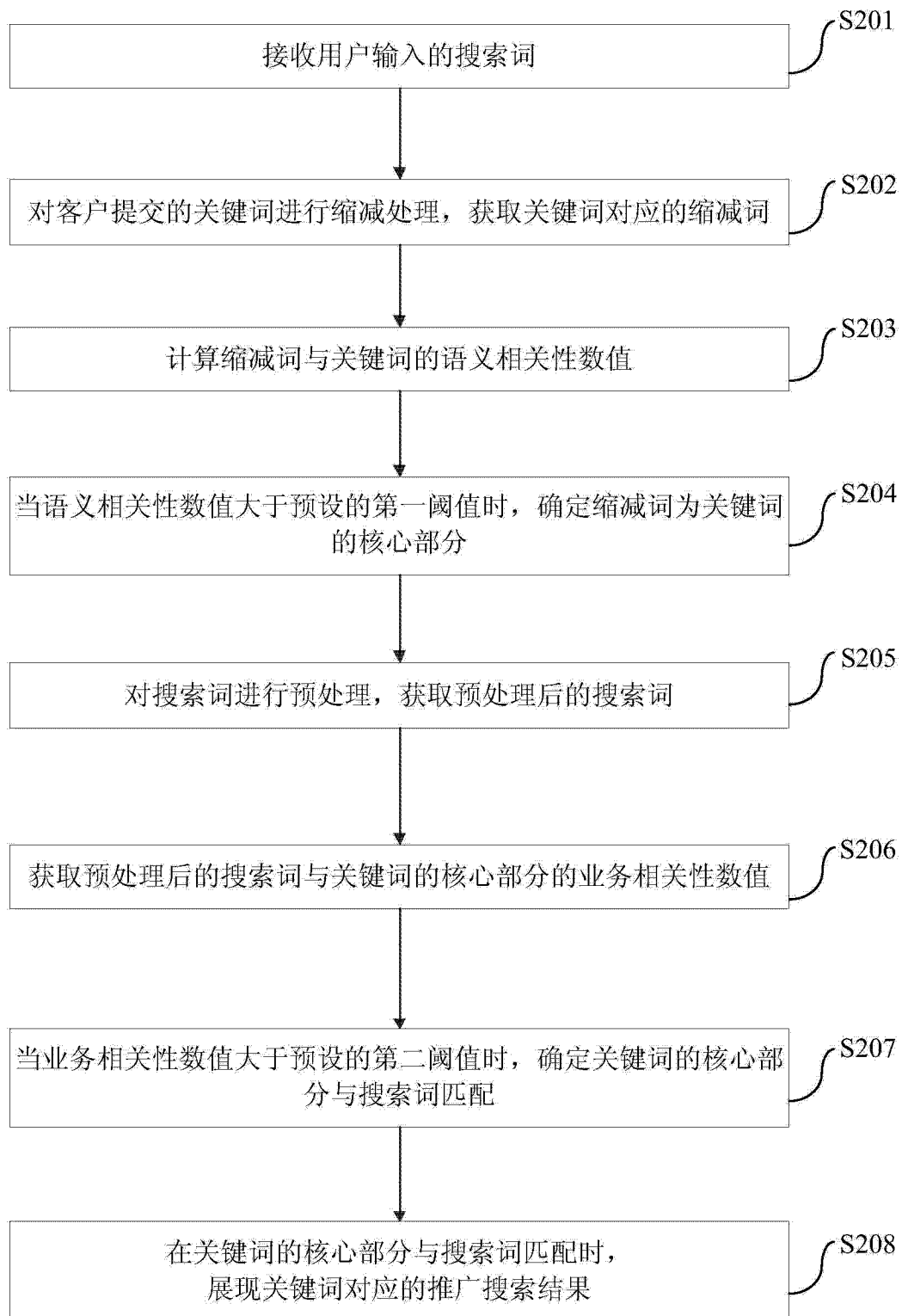


图 2

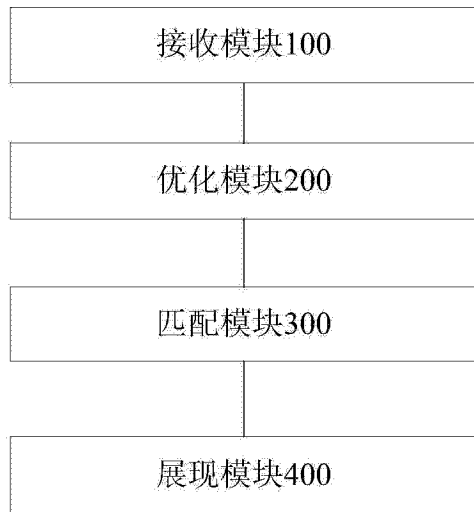


图 3

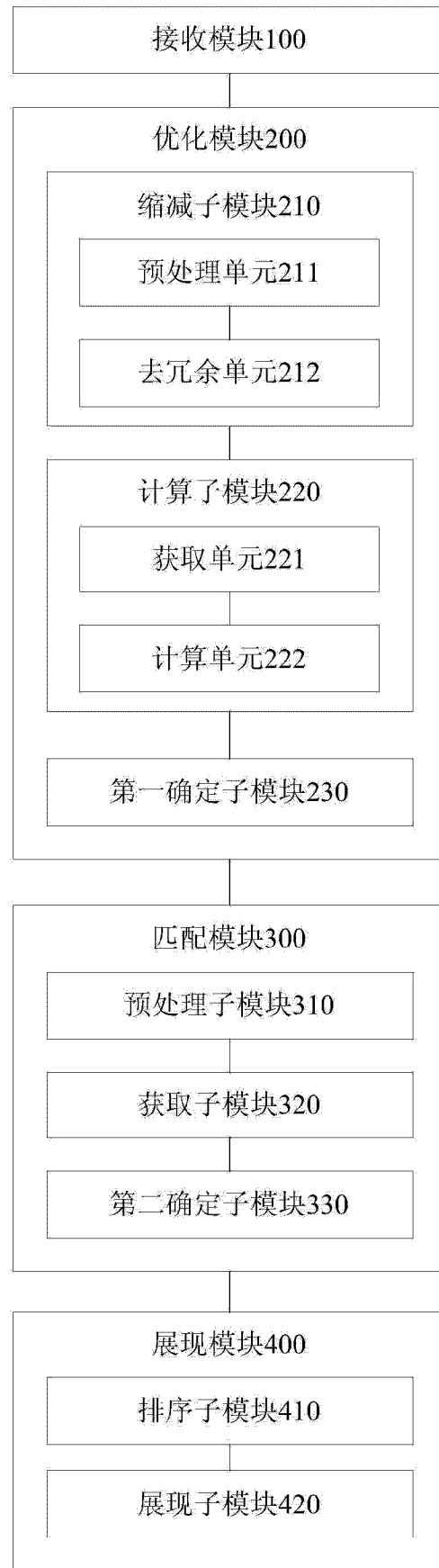


图 4