



(12) 发明专利申请

(10) 申请公布号 CN 115186065 A

(43) 申请公布日 2022. 10. 14

(21) 申请号 202210842766.9

(22) 申请日 2022.07.18

(71) 申请人 中国银行股份有限公司
地址 100818 北京市西城区复兴门内大街1号

(72) 发明人 蔡红镀

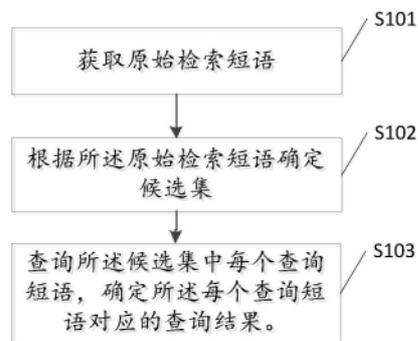
(74) 专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 马小青

(51) Int. Cl.
G06F 16/33 (2019.01)
G06F 16/35 (2019.01)
G06F 16/383 (2019.01)
G06F 40/194 (2020.01)
G06F 40/30 (2020.01)

权利要求书2页 说明书7页 附图3页

(54) 发明名称
一种目标词语的检索方法及装置

(57) 摘要
本申请公开了一种目标词语的检索方法及装置,可应用于人工智能领域。通过获取原始检索短语;根据所述原始检索短语确定候选集;查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。该方法通过引入潜在语义来提取检索库语料中所包含的潜在语义信息,基于原始查询构建语义相似的候选集查询,以解决现有全文搜索缺少语义匹配能力的问题,提升全文搜索的智能程度和用户体验。



1. 一种目标词语的检索方法,其特征在于,所述方法包括:
 - 获取原始检索短语;
 - 根据所述原始检索短语确定候选集;所述候选集包括原始检索短语和多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语
 - 查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。
2. 根据权利要求1所述的目标词语的检索方法,其特征在于,所述根据所述原始检索短语确定候选集,包括:
 - 获取潜在语义计算模型;
 - 通过所述潜在语义计算模型与第一规则确定多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;所述第一规则用于确定所述第一检索短语的个数;
 - 将所述原始检索短语与所述多个第一检索短语合并形成候选集。
3. 根据权利要求1所述的目标词语的检索方法,其特征在于,所述确定所述每个查询短语对应的查询结果,包括:
 - 根据所述查询短语确定文本搜索记录,所述文本搜索记录包括文本相关词语;
 - 根据第二规则确定所述查询结果,所述第二规则用于确定所述查询结果中所述文本相关词语的个数。
4. 根据权利要求1所述的目标词语的检索方法,其特征在于,所述确定所述每个查询短语对应的查询结果之后,还包括:
 - 合并所述每个查询短语对应的查询结果,将合并结果作为最终查询结果集合。
5. 根据权利要求2所述的目标词语的检索方法,其特征在于,所述潜在语义计算模型为隐性语义分析模型或词向量模型。
6. 一种目标词语的检索装置,其特征在于,所述装置包括:
 - 原始检索短语获取模块,用于获取原始检索短语;
 - 候选集确定模块,用于根据所述原始检索短语确定候选集;所述候选集包括原始检索短语和多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;
 - 查询结果确定模块,用于查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。
7. 根据权利要求6所述的装置,其特征在于,所述候选集确定模块包括:
 - 计算模型获取模块,用于获取潜在语义计算模型;
 - 第一检索短语确定模块,用于通过所述潜在语义计算模型与第一规则确定多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;所述第一规则用于确定所述第一检索短语的个数;
 - 候选集形成模块,用于将所述原始检索短语与所述多个第一检索短语合并形成候选集。
8. 根据权利要求6所述的装置,其特征在于,所述查询结果确定模块包括:
 - 文本搜索记录确定模块,用于根据所述查询短语确定文本搜索记录,所述文本搜索记录包括文本相关词语;
 - 查询结果确定模块,用于根据第二规则确定所述查询结果,所述第二规则用于确定所述查询结果中所述文本相关词语的个数。

9. 根据权利要求6所述的装置,其特征在于,所述装置还包括:
合并模块,用于合并所述每个查询短语对应的查询结果,将合并结果作为最终查询结果集合。
10. 根据权利要求7所述的装置,其特征在于,所述潜在语义计算模型为隐性语义分析模型或词向量模型。

一种目标词语的检索方法及装置

技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种目标词语的检索方法及装置。

背景技术

[0002] 全文检索是将存储于数据库中整本书、整篇文章中的任意内容信息查找出来的检索。它可以根据需要获得全文中有关章、节、段、句、词等信息,也就是说类似于给整本书的每个字词添加一个标签,也可以进行各种统计和分析。目前常用的全文搜索引擎如ES检索底层采用TFIDF(词频-逆向文件频率)算法计算相关度,该类方法相对具有局限性,TFIDF算法不能挖掘出词汇间深层次的语义关系,导致传统ES搜索引擎无法处理一义多词的情况,停留在低级别的关键词检索层面,无法提供给用户语义层面的检索。例如:用户搜索“automobile”,即汽车,传统全文搜索仅仅会返回包含“automobile”单词的记录,而实际上包含“car”单词的记录也可能是用户所需要的。

[0003] 也就是说,在当前这种目标词语的检索方法中,搜索结果通常会拘泥于用户所输入请求语句的字面本身,无法深层次捕捉到用户所输入语句后面的真正意图。这样的词语的检索方法查全性较差、差准性较低。

发明内容

[0004] 有鉴于此,本申请实施例提供了一种目标词语的检索方法及装置,旨在实现目标词语的精准全文检索。

[0005] 第一方面,本申请实施例提供了一种目标词语的检索方法及装置,所述方法包括:

[0006] 获取原始检索短语;

[0007] 根据所述原始检索短语确定候选集;所述候选集包括原始检索短语和多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语

[0008] 查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。

[0009] 可选的,所述根据所述原始检索短语确定候选集,包括:

[0010] 获取潜在语义计算模型;

[0011] 通过所述潜在语义计算模型与第一规则确定多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;所述第一规则用于确定所述第一检索短语的个数;

[0012] 将所述原始检索短语与所述多个第一检索短语合并形成候选集。

[0013] 可选的,所述确定所述每个查询短语对应的查询结果,包括:

[0014] 根据所述查询短语确定文本搜索记录,所述文本搜索记录包括文本相关词语;

[0015] 根据第二规则确定所述查询结果,所述第二规则用于确定所述查询结果中所述文本相关词语的个数。

[0016] 可选的,所述确定所述每个查询短语对应的查询结果之后,还包括:

[0017] 合并所述每个查询短语对应的查询结果,将合并结果作为最终查询结果集合。

- [0018] 可选的,所述潜在语义计算模型为隐性语义分析模型或词向量模型。
- [0019] 第二方面,本申请实施例提供了一种目标词语的检索装置,所述装置包括:
- [0020] 原始检索短语获取模块,用于获取原始检索短语;
- [0021] 候选集确定模块,用于根据所述原始检索短语确定候选集;所述候选集包括原始检索短语和多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;
- [0022] 查询结果确定模块,用于查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。
- [0023] 可选的,所述候选集确定模块包括:
- [0024] 计算模型获取模块,用于获取潜在语义计算模型;
- [0025] 第一检索短语确定模块,用于通过所述潜在语义计算模型与第一规则确定多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;所述第一规则用于确定所述第一检索短语的个数;
- [0026] 候选集形成模块,用于将所述原始检索短语与所述多个第一检索短语合并形成候选集。
- [0027] 可选的,所述查询结果确定模块包括:
- [0028] 文本搜索记录确定模块,用于根据所述查询短语确定文本搜索记录,所述文本搜索记录包括文本相关词语;
- [0029] 查询结果确定模块,用于根据第二规则确定所述查询结果,所述第二规则用于确定所述查询结果中所述文本相关词语的个数。
- [0030] 可选的,所述装置还包括:
- [0031] 合并模块,用于合并所述每个查询短语对应的查询结果,将合并结果作为最终查询结果集合。
- [0032] 可选的,所述潜在语义计算模型为隐性语义分析模型或词向量模型。
- [0033] 本申请实施例提供了一种目标词语的检索方法及装置。在执行所述方法时,获取原始检索短语;根据所述原始检索短语确定候选集;查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。由此,当用户输入目标检索短语后,系统分析目标检索短语获得同义短语,以原始待检索文本和多条语义相似文本记录共同作为检索条件,获得多个检索结果。这样,达到了对目标词语的全面检索效果。如此,检索结果综合了关键词检索和语义检索,提升现有全文检索的查全率和查准率。

附图说明

- [0034] 为更清楚地说明本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申请的一些实施例,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。
- [0035] 图1为本申请实施例提供的目标词语的检索的一种方法流程图;
- [0036] 图2为本申请实施例提供的目标词语的检索的一种方法流程图;
- [0037] 图3为本申请实施例提供的目标词语的检索的一种结构示意图。

具体实施方式

[0038] 为了使本技术领域的人员更好地理解本申请方案,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0039] 正如前文所述,当前目前常用的全文搜索引擎如ES检索底层采用TFIDF算法计算相关度。但是,发明人经过研究发现,该类方法相对具有局限性,TFIDF算法不能挖掘出词汇间深层次的语义关系,导致该种检索方法无法深层次捕捉到用户所输入语句后面的真正意图,该方法具有查全性较差、差准性较低的缺点。

[0040] 为了解决这一问题,在本申请实施例提供了一种目标词语的检索方法及装置,在执行所述方法时,获取原始检索短语;根据所述原始检索短语确定候选集;查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。由此,当用户输入目标检索短语后,系统分析目标检索短语获得同义短语,以原始待检索文本和多条语义相似文本记录共同作为检索条件,获得多个检索结果。这样,达到了对目标词语的全面检索效果。如此,检索结果综合了关键词检索和语义检索,提升现有全文检索的查全率和查准率。

[0041] 本申请实施例提供的方法由搜索引擎和后台服务器执行,例如后台服务器中包括具有检索功能和整合功能的检索系统。检索系统获取检索库中的语义计算模型后,对原始检索短语进行分析获取语义相近的短语,系统将查询对象输入搜索引擎例如ES检索,根据相关度形成查询结果。所述后台服务器可以是一台服务器设备,也可以是由多台服务器组成的服务器集群。

[0042] 以下通过一个实施例,对本申请提供的目标词语的检索方法进行说明。请参考图1,图1为本申请实施例所提供的目标词语的检索方法的一种方法流程图,包括:

[0043] S101:获取原始检索短语。

[0044] 原始检索短语为初始检索目标短语。在具体应用场景中,原始检索短语可以由用户输入的,也可以是系统根据查询需求设置的。

[0045] S102:根据所述原始检索短语确定候选集。

[0046] 所述候选集包括原始检索短语和多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语。

[0047] 假设查询原始检索短语为“大数据调研”,通过潜在语义模型计算待检索短语在检索库中语义最相似的前2条记录为“调研”、“学习”,则这两条记录为第一检索短语。系统可以将查询获得的原始检索短语和多条第一检索短语纳入候选集。在实际应用场景中,系统可以根据不同第一检索短语与原始检索短语的不同相关度,各个第一检索短语设置不同的标志,可以用于后续过程中对相似度不同的检索结果进行区分。

[0048] 关于如何根据原始检索短语确定候选集,具体参见后文,在此不做赘述。

[0049] S103:查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。

[0050] 查询文本为候选集中逐个进行查询的短语。从候选集中依次取出原始检索短语和多个第一检索短语,逐个作为查询短语。确定每个查询短语对应的查询结果。

[0051] 在实际应用场景中,系统可以将每条查询短语对应的查询结果合并,形成最终查询结果集合。

[0052] 下面对本申请实施例提供的目标词语的检索方法进行详细介绍。参见图2所示，图2为本申请实施例通过的目标词语的检索的另一种流程示意图。其具体过程如下：

[0053] S201:获取原始检索短语。

[0054] 系统获取需要进行检索的原始检索短语。

[0055] S202:获取潜在语义计算模型。

[0056] 获取检索库中语料离线计算潜在语义模型，该模型可以是LSA模型也可以是Word2vec模型，这里不做限制，下面步骤实现用LSA举例，具体包括：

[0057] 分析文档集合，建立词汇-文本矩阵A。假设A是一个 $m \times n$ 的文本数据矩阵($n \ll m$)，表示该语料包含 m 个单词， n 篇文档。

[0058] 对词汇-文本矩阵进行奇异值分解，对SVD分解后的矩阵进行降维，使用降维后的矩阵构建潜在语义空间LSA模型，公式为：

$$[0059] \quad A_{m \times n} = U_{m \times k} \Sigma_{k \times k} V_{k \times n}^T \approx U_{m \times r} \Sigma_{r \times r} V_{r \times n}^T$$

[0060] 式中， $A_{m \times n}$ 为 $m \times n$ 文本数据矩阵，该式将大矩阵A进行截断奇异值分解，分解为3个矩阵的乘积矩阵， $U_{m \times k}$ 为单词-话题矩阵， $\Sigma_{k \times k} V_{k \times n}^T$ 为话题文本矩阵，A可以分解出 k 个特征值， k 在此处可以代指主题数，我们选取排序后其中较大的 r 个特征值， r 的值可以根据下列公式计算得出：

$$[0061] \quad \frac{P_r}{P} \geq 95\%$$

[0062] 式中 P_r 为对角矩阵的前 r 个较大特征值的平方和， P 为对角矩阵所有特征值的平方和，计算得到的 r 能够拥有原矩阵95%以上的信息量，且 r 远小于 k 。

[0063] 这样 $U_{m \times r} \Sigma_{r \times r} V_{r \times n}^T$ 就可以近似表示矩阵A。 U 是单词-话题矩阵，每一列代表一个潜语义，这个潜语义的意义由 m 个单词按不同权重组合而成，其行代表词语，其列代表文档。一般情况下，词-文档矩阵的元素是该词在文档中的出现次数。因为 U 中每一列相互独立，所以 r 个潜语义构成了一个语义空间，矩阵 U 中的每一列表示一个关键词，数值越大则越相关，因此，通过 $U_{m \times r}$ 可以看到词与词义之间的相关性。矩阵 V 中的每一行表示一类主题，其中的每个非零元素表示一个主题与一个文档的相关性，从 $V_{r \times n}^T$ 可以看出文本与主题的相关性。而 ΣV^T 是话题-文档矩阵， ΣV^T 中每一列代表一篇文档，将该文档被映射到了语义空间， Σ 中每一个奇异值指示了该潜语义的重要度，矩阵 Σ 表示文章主题和关键词之间的相关性。

[0064] S203:通过所述潜在语义计算模型与第一规则确定多个第一检索短语。

[0065] 其中，第一规则用于确定所述第一检索短语的个数。在具体应用场景中，第一规则可以由用户自行设置，也可以是由系统根据查询需求设置。

[0066] 在一些可能的实现方式中，用户可以根据个人搜索需求，对系统提供的语义相近的搜索记录进行选取或剔除，也可以由用户根据需求选取不同的第一检索短语和原始检索短语形成不同的组合。例如当用户删除语义第二相近的检索短语，则语义第三相近的检索短语可以主动补位。第一检索短语也可以由用户自行设置，选取系统检索获得的其他语义相近的检索短语或自行输入第一检索短语内容。

[0067] 例如，当第一规则指示第一检索短语个数为2，则在实际应用场景中，系统通过潜在语义模型计算待检索短语在检索库中语义最相似的前2条记录。假设查询为：query=“大数据调研”，通过潜在语义模型计算待检索短语在检索库中语义最相似的前2条记录为：

similar_top_2 = {“hadoop调研”、“spark学习”}。其中，“调研”和“学习”为当前第一规则对应的第一检索短语。

[0068] 其中,Hadoop和Spark两者为不同第一检索短语的代表标志。在实际应用过程中,系统可以根据不同的代表标志相似度不同的第一检索短语进行区分,在一些可能的实现方式里,系统可以对第一检索短语的代表标志进行设置和适应性修改。

[0069] 在实际应用过程中,Hadoop和Spark两者都是大数据框架,Spark是专为大规模数据处理而设计的快速通用的计算引擎。Hadoop是一个由Apache基金会所开发的分布式系统基础架构,Hadoop以一种可靠、高效、可伸缩的方式进行数据处理。在应用过程中,可以根据这两种大数据架框对系统检索得到的数据进行进一步的数据处理。

[0070] 作为进一步优化,S203中通过潜在语义模型计算待检索短语在检索库中语义最相似的前N条记录,具体包括:

[0071] 对于给定的查询,我们根据这个查询中包含的单词 A_q 构造一个伪文档: $V_q = A_q U \Sigma^{-1}$,然后该伪文档和V中的每一列计算余弦相似度来得到和给定查询最相似的N个文档。假设V中第t列对应的文本向量为 V_t ,则伪文档向量 V_q 和 V_t 之间的余弦相似度计算公式为:

$$[0072] \quad \cos(V_q, V_t) = \frac{V_q \cdot V_t}{\|V_q\| \times \|V_t\|}$$

[0073] 式中 V_q 与 V_t 为查询文本与语义空间矩阵中第t列对应的文本的向量表示, $\|V_q\|$ 与 $\|V_t\|$ 分别是向量 V_q 与 V_t 的模, $\cos(V_q, V_t)$ 为文本向量与为文档向量之间的余弦相似度。

[0074] S204:将所述原始检索短语与所述多个第一检索短语合并形成候选集。

[0075] 将原始检索短语与第一规则对应的多个第一检索短语归并形成候选集,例如,将原始检索短语“大数据调研”与第一检索短语“hadoop调研”“spark学习”合并,得到candidate_list = {“大数据调研”,“hadoop调研”、“spark学习”}。

[0076] S205:查询所述候选集中每个查询短语,根据所述查询短语确定文本搜索记录。

[0077] 其中,文本搜索记录包括多个文本相关词语,文本相关词语为查询出的与查询短语具有相关性的词语。文本搜索记录为多个相关性词语形成的集合。

[0078] 具体的,系统从候选集中依次取出需要查询的短语,根据查询短语确定出当前短语对应的文本搜索记录,例如,查询文本为原始检索短语,即query1 = “大数据调研”,当前查询短语对应的文本搜索记录可以为“大数据发展现状调研”、“研究大数据相关组件”、“大数据与人工智能关系”、“大数据发展前景探究”“大数据的应用拓展”,文本搜索记录为基于当前查询文本的全文检索结果,也就是说,若全文中由N个与当前查询短语直接相关的内容,则文本搜索记录可以为N条。

[0079] 在一些可能的实现方式中,系统基于与查询短语的相关性高低可以对各个相关性词语进行排序。即在系统获取文本相关短语的过程中,根据相关度对获取到的若干文本相关短语进行排序,形成具有顺序性的文本搜索记录。因此,在当前文本搜索记录中,多个文本相关词语之间存在相关度顺序关系,以查询短语大数据调研为相关度判断标准,大数据发展现状调研相关度高于研究大数据相关组件,相关度向后逐渐递减。

[0080] 在一些可能的实现方式中,系统仅根据查询短语检索全文,未在检索过程中对获得的相关性的词语排序。因此,形成的文本搜索记录中各个文本相关词语之间没有相关性高低顺序关系。关于根据相关度对文本搜索记录中文本相关词语进行提取,系统在获取到

第二规则对应的个数请求后,可以对文本搜索记录中的文本相关词排序,也可以设置相关度阈值直接对文本搜索记录中的多个文本相关词进行筛选,进而获得第二规则对应个数的文本相关词。即系统从无顺序关系的文本搜索记录集合中选取若干个相关度达标的文本相关短语。

[0081] S206:根据第二规则确定所述查询结果。

[0082] 第二规则用于确定所述查询结果中所述文本相关词语的个数。在具体应用场景中,第二规则可以由用户自行设置,也可以是由系统根据查询需求设置。

[0083] 例如,当第二规则指示文本相关短语个数为3,则在实际应用场景中,系统取相关度前3的文本记录形成查询结果。假设当前查询短语对应的文本搜索记录为“大数据发展现状调研”、“研究大数据相关组件”、“大数据与人工智能关系”、“大数据发展前景探究”“大数据的应用拓展”,根据S205步骤中提及的文本相关短语选取规则,选取相关度前3的文本记录形成查询结果,则对应的结果集 $result1 = \{$ ”大数据发展现状调研“、“研究大数据相关组件”、“大数据与人工智能关系” $\}$ 。

[0084] S207:合并所述每个查询短语对应的查询结果,将合并结果作为最终查询结果集合。

[0085] 根据上述S206步骤中,确定候选集中的若干查询文本对应的查询结果。

[0086] 例如,从候选集中依次取出各个查询文本输入全文搜索引擎例如ES检索,取相关度前3的文本记录形成查询结果。查询文本 $query1 =$ “大数据调研”, $query2 =$ “hadoop调研”, $query3 =$ “spark学习”,对应的结果集 $result1 = \{$ “大数据发展现状调研”、“研究大数据相关组件”、“大数据与人工智能关系” $\}$, $result2 = \{$ “hadoop调研”、“hadoop技术调研”、“hadoop快速入门” $\}$, $result3 = \{$ ”spark学习“、“spark学习笔记”、“spark基础教程” $\}$ 。在当前步骤中,合并每条查询短语对应的查询结果构成检索结果集合为: $result_list = \{$ ”大数据发展现状调研“、“研究大数据相关组件”、“大数据与人工智能关系”、“hadoop调研“、“hadoop技术调研”、“hadoop快速入门”、“spark学习“、“spark学习笔记”、“spark基础教程” $\}$ 。

[0087] 在实际应用场景中,可以在用户端搜索界面对查询结果分类显示,例如,可以将原始检索短语以及相似度不同的第一检索短语进行区分,设置原始检索短语的查询结果位于第一行,而后,第一检索短语对应的查询结果随行数相似度递减。或者,系统可以对不同的短语设置不同的颜色,进而达到在显示界面的区分效果。

[0088] 以上为本申请实施例提供一种基于潜在语义分析的检索方法的一些具体实现方式,基于此,本申请还提供了对应的装置。下面将从功能模块化的角度对本申请实施例提供的装置进行介绍。

[0089] 请参考图3,图3为本申请实施例所提供的一种目标词语的检索装置的结构示意图。

[0090] 本实施例中,该装置可以包括:

[0091] 原始检索短语获取模块300,用于获取原始检索短语;

[0092] 候选集确定模块301,用于根据所述原始检索短语确定候选集;所述候选集包括原始检索短语和多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;

[0093] 查询结果确定模块302,用于从目标文本中查询所述候选集中每个查询短语,确定所述每个查询短语对应的查询结果。

[0094] 可选的,所述候选集确定模块包括:

[0095] 计算模型获取模块,用于获取潜在语义计算模型;

[0096] 第一检索短语确定模块,用于通过所述潜在语义计算模型与第一规则确定多个第一检索短语,所述第一检索短语为与所述原始检索短语语义相近的短语;所述第一规则用于确定所述第一检索短语的个数;

[0097] 候选集形成模块,用于将所述原始检索短语与所述多个第一检索短语合并形成候选集。

[0098] 可选的,所述查询结果确定模块包括:

[0099] 文本搜索记录确定模块,用于根据所述查询短语确定文本搜索记录,所述文本搜索记录包括文本相关词语;

[0100] 查询结果确定模块,用于根据第二规则确定所述查询结果,所述第二规则用于确定所述查询结果中所述文本相关词语的个数。

[0101] 可选的,所述装置还包括:

[0102] 合并模块,用于合并所述每个查询短语对应的查询结果,将合并结果作为最终查询结果集合。

[0103] 可选的,所述装置还包括:

[0104] 所述潜在语义计算模型为隐性语义分析模型或词向量模型。

[0105] 需要说明的是,本发明提供的一种目标词语的检索方法及装置可用于人工智能领域。上述仅为示例,并不对本发明提供的一种目标词语的检索方法及装置的应用领域进行限定。

[0106] 以上对本申请所提供的一种目标词语的检索方法及装置进行了详细介绍。说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请原理的前提下,还可以对本申请进行若干改进和修饰,这些改进和修饰也落入本申请权利要求的保护范围内。

[0107] 还需要说明的是,在本说明书中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,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、物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、物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,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0108] 以上所述仅为本申请的较佳实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

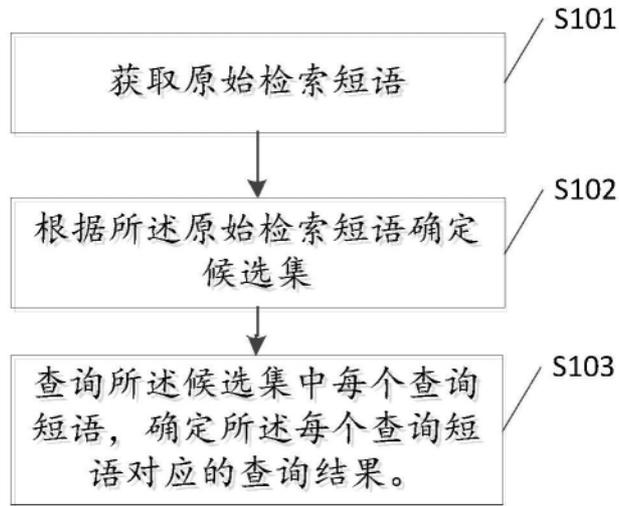


图1

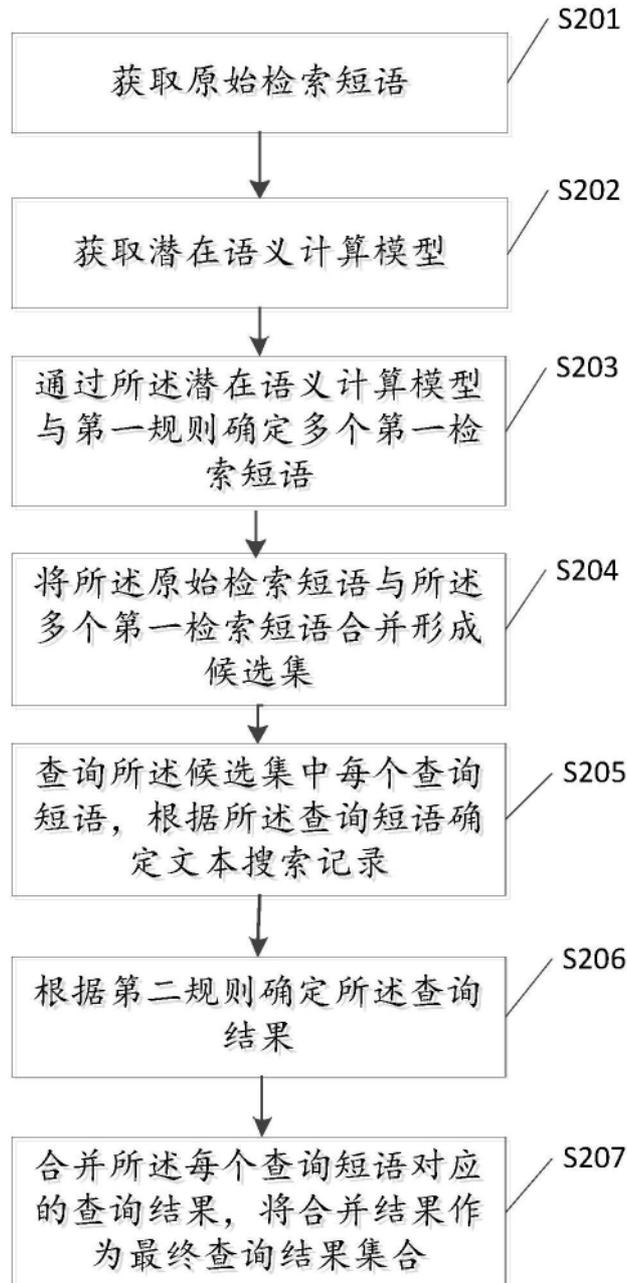


图2

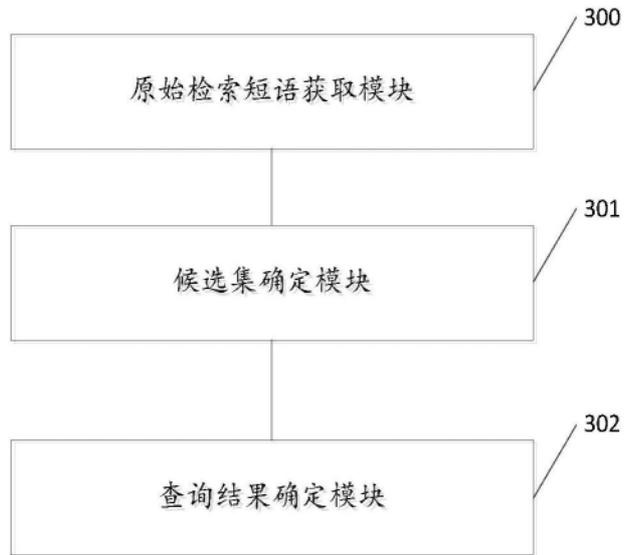


图3