



# (12) 发明专利

(10) 授权公告号 CN 111061835 B

(45) 授权公告日 2023. 09. 22

(21) 申请号 201911301718.3

G06N 3/08 (2023.01)

(22) 申请日 2019.12.17

(56) 对比文件

(65) 同一申请的已公布的文献号

申请公布号 CN 111061835 A

CN 109256216 A, 2019.01.22

WO 2015180622 A1, 2015.12.03

CN 105589972 A, 2016.05.18

(43) 申请公布日 2020.04.24

KR 20100129016 A, 2010.12.08

(73) 专利权人 医渡云(北京)技术有限公司

CN 106407280 A, 2017.02.15

地址 100191 北京市海淀区花园北路35号9

CN 109522393 A, 2019.03.26

号楼8层801

CN 101246492 A, 2008.08.20

(72) 发明人 张子锐

CN 107958406 A, 2018.04.24

US 2018107933 A1, 2018.04.19

(74) 专利代理机构 北京律智知识产权代理有限公司

公司 11438

CN 107657062 A, 2018.02.02

专利代理师 袁礼君 阚梓瑄

CN 108304444 A, 2018.07.20

冯勇等.融合TF-IDF和LDA的中文FastText短文本分类方法.《应用科学学报》.2019, (03), 全文.

(51) Int. Cl.

G06F 16/33 (2019.01)

G06F 16/332 (2019.01)

G16H 50/70 (2018.01)

审查员 李利华

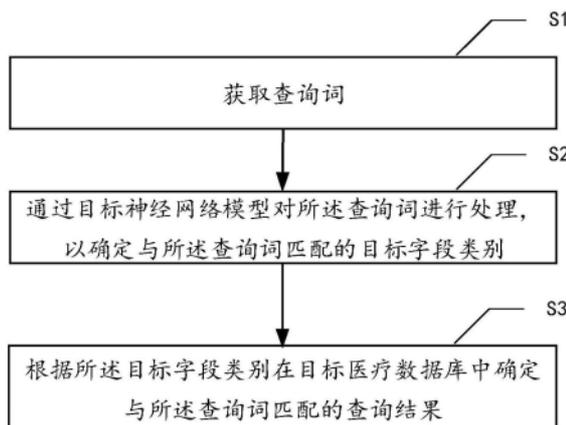
权利要求书2页 说明书12页 附图6页

(54) 发明名称

查询方法及装置、电子设备和计算机可读存储介质

(57) 摘要

本公开提供一种查询方法、装置以及电子设备和计算机可读存储介质,该方法包括:获取查询词;根据所述查询词确定目标字段类别;根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。本公开实施例提供的技术方案根据可以根据查询词快速、准确的确定与查询词相关的查询结果。



1. 一种查询方法,其特征在于,包括:

获取查询词;

通过目标神经网络模型对所述查询词进行处理,以确定与所述查询词匹配的目标字段类别,其中所述目标字段类别是电子病历中字段的名称,所述目标字段类别包括姓名、既往史、过敏史、诊断、门诊医嘱、手术、物理检查、实验室检查、药品名称、症状以及科室;

确定所述查询词的同义词和子类词,其中所述查询词的子类词是所述查询词的子类别;

在所述目标字段类别对应的内容中查询出与所述查询词、所述查询词的同义词以及所述查询词的子类词匹配的查询结果;

根据与所述查询词、所述查询词的同义词以及所述查询词的子类词匹配的查询结果,在目标医疗数据库中确定与所述查询词匹配的查询结果;

其中,所述目标神经网络模型是通过以下方法训练获得的:

从电子病历中抽取待检索的字段作为目标字段样本,所述目标字段样本包括字段名称和字段值,其中所述字段名称作为所述目标字段样本对应的标签;

通过所述目标字段的字段值和所述目标字段的字段名称对所述目标神经网络进行训练。

2. 根据权利要求1所述方法,其特征在于,通过所述目标字段的字段值和所述目标字段的字段名称对所述目标神经网络进行训练,包括:

根据N-gram方法对所述目标字段样本的值进行向量化处理,以获得目标向量;

根据所述目标字段样本的目标向量以及所述目标字段样本的字段名称训练所述目标神经网络模型。

3. 根据权利要求1所述方法,其特征在于,所述方法还包括:

对所述查询词进行分词处理,以获得所述查询词的分词结果;

对所述查询词的相关词进行分词处理,以获得所述相关词的分词结果;

根据所述查询词的分词结果和所述相关词的分词结果在所述目标字段类别中确定与所述查询词匹配的信息。

4. 根据权利要求1所述方法,其特征在于,所述目标字段类别包括第一字段类别和第二字段类别;其中,在所述目标字段类别对应的内容中查询出与所述查询词匹配的查询结果包括:

根据所述第一字段类别在所述目标医疗数据库中确定与所述查询词匹配的第一结果;

根据所述第二字段类别在所述目标医疗数据库中确定与所述查询词匹配的第二结果;

对所述第一结果和所述第二结果进行去重处理,以确定与所述查询词匹配的查询结果。

5. 一种查询装置,其特征在于,包括:

查询词获取模块,配置为获取查询词;

目标字段类别确定模块,配置为通过目标神经网络模型对所述查询词进行处理,以确定与所述查询词匹配的目标字段类别,其中所述目标字段类别是电子病历中字段的名称,所述目标字段类别包括姓名、既往史、过敏史、诊断、门诊医嘱、手术、物理检查、实验室检查、药品名称、症状以及科室;

查询结果获取模块,配置为确定所述查询词的同义词和子类词,其中所述查询词的子类词是所述查询词的子类别;在所述目标字段类别对应的内容中查询出与所述查询词、所述查询词的同义词以及所述查询词的子类词匹配的查询结果;根据与所述查询词、所述查询词的同义词以及所述查询词的子类词匹配的查询结果,在目标医疗数据库中确定与所述查询词匹配的查询结果;其中,所述目标神经网络模型是通过以下方法训练获得的:从电子病历中抽取待检索的字段作为目标字段样本,所述目标字段样本包括字段名称和字段值,其中所述字段名称作为所述目标字段样本对应的标签;通过所述目标字段的字段值和所述目标字段的字段名称对所述目标神经网络进行训练。

6. 一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-4中任一项所述的方法。

7. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行时实现如权利要求1-4中任一项所述的方法。

## 查询方法及装置、电子设备和计算机可读存储介质

### 技术领域

[0001] 本公开涉及信息分析及处理技术领域,尤其涉及一种查询方法及装置、电子设备和计算机可读存储介质。

### 背景技术

[0002] 随着医疗信息化的普及,各个医院都逐渐建立了比较成熟的医疗信息系统,各个医院也积累了大量的诊断记录、入院记录、用药处方等电子病历。这些数据对医学科学研究有很大的潜在价值。而在医疗研究中通常需要检索相关的数据。例如,医疗研究者有时需要检索与糖尿病有关的信息。而与糖尿病有关的信息可能会包括患有该糖尿病的病人,治疗糖尿病做的特定手术名、物理检查、实验室检查等,以及治疗该糖尿病使用的药品(包括药品通用名称、化学名称等)等。

[0003] 然而,当医疗研究者在医疗数据库中进行检索时,经常会检索出一些与医疗研究者的检索目的无关的信息。例如,当医疗研究者想要通过检索词“高血压”检索出被诊断为“高血压”疾病的患者关的病历时,通常还会检索出医院内科室名称为“高血压病房”的病历。很明显,直接在医疗数据库中进行大范围的检索不仅检索效率低,而且还会命中一些无关的医疗数据。

[0004] 因此,如何在复杂的医疗信息中高效、精确地查找医疗研究者想要的医疗数据成为了一个难题。

[0005] 需要说明的是,在上述背景技术部分公开的信息仅用于加强对本公开的背景的理解,因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

### 发明内容

[0006] 本公开实施例提供一种查询方法及装置、电子设备和计算机可读存储介质,能够准确的在目标医疗数据库中找到与查询词匹配的医疗类别,并根据医疗类别确定与查询词匹配的查询结果。

[0007] 本公开的其他特性和优点将通过下面的详细描述变得显然,或部分地通过本公开的实践而习得。

[0008] 本公开实施例提出一种查询方法,该方法包括:获取查询词;根据所述查询词确定目标字段类别;根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。

[0009] 在一些实施例中,根据所述查询词确定目标字段类别,包括:通过目标神经网络模型对所述查询词进行处理以获得所述目标字段类别。

[0010] 在一些实施例中,所述查询方法还包括:获取目标字段样本,所述目标字段样本包括医疗类别信息,所述目标字段样本取自于医疗病历;根据所述目标字段样本以及所述目标字段样本的医疗类别信息训练所述目标神经网络模型。

[0011] 在一些实施例中,根据所述目标字段样本以及所述目标字段样本的医疗类别信息

训练所述目标神经网络模型,包括:根据N-gram方法对所述目标字段样本进行向量化处理,以获得目标向量;根据所述目标字段样本的目标向量以及所述目标字段样本的类别信息训练所述目标神经网络模型。

[0012] 在一些实施例中,根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果包括:确定所述查询词的相关词;根据所述查询词以及所述查询词的相关词在目标字段类别中确定与所述查询词匹配的信息。

[0013] 在一些实施例中,所述查询词的相关词包括所述查询词的同义词以及所述查询词的子类词;其中,根据所述查询词以及所述查询词的相关词在目标字段类别中确定与所述查询词匹配的信息,包括:根据所述查询词、所述查询词的同义词以及所述查询词的子类词在目标字段类别中确定与所述查询词匹配的信息。

[0014] 在一些实施例中,根据所述查询词以及所述查询词的相关词在目标字段类别中确定与所述查询词匹配的信息,包括:对所述查询词进行分词处理,以获得所述查询词的分词结果;对所述查询词的相关词进行分词处理,以获得所述相关词的分词结果;根据所述查询词的分词结果和所述相关词的分词结果在所述目标字段类别中确定与所述查询词匹配的信息。

[0015] 在一些实施例中,所述目标字段类别包括第一字段类别和第二字段类别;其中,根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果,包括:根据所述第一字段类别在所述目标医疗数据库中确定与所述查询词匹配的第一结果;根据所述第二字段类别在所述目标医疗数据库中确定与所述查询词匹配的第二结果;对所述第一结果和所述第二结果进行去重处理,以确定与所述查询词匹配的查询结果。

[0016] 本公开实施例提供了一种查询装置,其特征在于,包括:查询词获取模块、目标字段类别确定模块以及查询结果获取模块。

[0017] 其中,所述查询词获取模块可以配置为获取查询词;所述目标字段类别确定模块可以配置为根据所述查询词确定目标字段类别;所述查询结果获取模块可以配置为根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。

[0018] 本公开实施例提出一种电子设备,该电子设备包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现上述任一项所述的查询方法。

[0019] 本公开实施例提出一种计算机可读存储介质,其上存储有计算机程序,所述程序被处理器执行时实现如上述任一项所述的查询方法。

[0020] 本公开某些实施例提供的查询方法、装置及电子设备和计算机可读存储介质,通过查询词确定了目标字段类别,并进一步的在目标字段类别中确定了与查询词相关的查询信息。该方法一方面相比于在包括所有医疗数据的数据库中进行信息查询的方法,可以极大地提高查询效率,减少不必要的资源浪费;另一方面可以精准定位至目标字段类别中进行信息的查询,以准确获取与查询词相关的查询信息,提供查询命中率。例如,当目标用户输入“高血压”时,根据本方案提供的查询方法可能会查到与治疗高血压相关的治疗信息,而不会是“高血压病房”等无关信息。

[0021] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性的,并不能限制本公开。

## 附图说明

[0022] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。下面描述的附图仅仅是本公开的一些实施例,对于本领域的普通技术人员来讲,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。

[0023] 图1示出了可以应用于本公开实施例的查询方法或查询装置的示例性系统架构的示意图。

[0024] 图2是根据一示例性实施例示出的一种应用于查询装置的计算机系统的结构示意图。

[0025] 图3是根据一示例性实施例示出的一种查询方法的流程图。

[0026] 图4是图3中步骤S2在一示例性实施例中的流程图。

[0027] 图5是图4中步骤S22在一示例性实施例中的流程图。

[0028] 图6是根据一示例性实施例示出的一种根据目标向量训练目标神经网络模型的示意图。

[0029] 图7是图3中步骤S3在一示例性实施例中的流程图。

[0030] 图8是图7中步骤S32在一示例性实施例中的流程图。

[0031] 图9是根据示例性实施例示出的另一种查询方法的流程图。

[0032] 图10是根据示例性实施例示出的一种查询装置的方框图。

## 具体实施方式

[0033] 现在将参考附图更全面地描述示例实施例。然而,示例实施例能够以多种形式实施,且不应被理解为限于在此阐述的实施例;相反,提供这些实施例使得本公开将全面和完整,并将示例实施例的构思全面地传达给本领域的技术人员。在图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。

[0034] 本公开所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中,提供许多具体细节从而给出对本公开的实施方式的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而省略特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知方法、装置、实现或者操作以避免模糊本公开的各方面。

[0035] 附图仅为本公开的示意性图解,图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。附图中所示的一些方框图不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0036] 附图中所示的流程图仅是示例性说明,不是必须包括所有的内容和步骤,也不是必须按所描述的顺序执行。例如,有的步骤还可以分解,而有的步骤可以合并或部分合并,因此实际执行的顺序有可能根据实际情况改变。

[0037] 本说明书中,用语“一个”、“一”、“该”、“所述”和“至少一个”用以表示存在一个或多个要素/组成部分/等;用语“包含”、“包括”和“具有”用以表示开放式的包括在内的意思

并且是指除了列出的要素/组成部分/等之外还可存在另外的要素/组成部分/等;用语“第一”、“第二”和“第三”等仅作为标记使用,不是对其对象的数量限制。

[0038] 下面结合附图对本公开实施例实施方式进行详细说明。

[0039] 图1示出了可以应用于本公开实施例的查询方法或查询装置的示例性系统架构的示意图。

[0040] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104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等。

[0041] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。其中,终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机、台式计算机、可穿戴设备、虚拟现实设备、智能家居等等。

[0042] 服务器105可以是提供各种服务的服务器,例如对用户利用终端设备101、102、103所进行操作的装置提供支持的后台管理服务器。后台管理服务器可以对接收到的请求等数据进行分析等处理,并将处理结果反馈给终端设备。

[0043] 服务器105可例如获取查询词;服务器105可例如通过目标神经网络模型对所述查询词进行处理,以确定与所述查询词匹配的目标字段类别;服务器105可例如根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。

[0044] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的,服务器105可以是一个实体的服务器,还可以为多个服务器组成,根据实际需要,可以具有任意数目的终端设备、网络和服务器。

[0045] 下面参考图2,其示出了适于用来实现本申请实施例的终端设备的计算机系统200的结构示意图。图2示出的终端设备仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0046] 如图2所示,计算机系统200包括中央处理单元(CPU)201,其可以根据存储在只读存储器(ROM)202中的程序或者从储存部分208加载到随机访问存储器(RAM)203中的程序而执行各种适当的动作和处理。在RAM 203中,还存储有系统200操作所需的各种程序和数据。CPU 201、ROM 202以及RAM 203通过总线204彼此相连。输入/输出(I/O)接口205也连接至总线204。

[0047] 以下部件连接至I/O接口205:包括键盘、鼠标等的输入部分206;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分207;包括硬盘等的储存部分208;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分209。通信部分209经由诸如因特网的网络执行通信处理。驱动器210也根据需要连接至I/O接口205。可拆卸介质211,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器210上,以便于从其上读出的计算机程序根据需要被安装入储存部分208。

[0048] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机软件产品,其包括承载在计算机可读存储介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分209从网络上被下载和安装,和/或从可拆卸

介质211被安装。在该计算机程序被中央处理单元(CPU)201执行时,执行本申请的系统中限定的上述功能。

[0049] 需要说明的是,本申请所示的计算机可读存储介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读存储介质,该计算机可读存储介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读存储介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0050] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0051] 描述于本申请实施例中所涉及到的模块和/或)单元和/或)子单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块和/或)单元和/或)子单元也可以设置在处理器中,例如,可以描述为:一种处理器包括发送单元、获取单元、确定单元和第一处理单元。其中,这些模块和/或)单元和/或)子单元的名称在某种情况下并不构成对该模块和/或)单元和/或)子单元本身的限定。

[0052] 作为另一方面,本申请还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读存储介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备可实现功能包括:获取查询词;通过目标神经网络模型对所述查询词进行处理,以确定与所述查询词匹配的目标字段类别;根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。

[0053] 在医疗技术领域,医疗数据逐日增多,已经形成了一个数据量巨大的医疗数据库。目标用户可以从所述医疗数据库中获取与目标对象(所述目标用户关心的内容,例如糖

尿病就可以是目标对象) 相关的信息。

[0054] 但是在实际查询中,目标用户从医疗数据库中获取的查询结果通常会存在大量的与查询词无关的内容或者不是目标用户想要的结果。例如,当目标用户希望查询与高血压治疗相关的信息时,可能会输入查询词“高血压”,然而目标医疗数据库有可能会给出“高血压病房”信息、“高血压科室信息”等信息。再例如,当目标用户想要检索神经外科的病历时,可能会输入“神经外科”等查询词,但是单纯的输入“神经外科”查询词可能会命中“急诊头部CT(Computed Tomography,即电子计算机断层扫描)示考虑蛛网膜下腔出血,建议转入综合医院神经外科继续治疗”等无关信息。很明显,直接在医疗数据库中进行信息的查询,获得的查询信息可能不是目标用户希望得到的信息,并且在医疗数据库中进行信息的匹配,可能会由于信息量过大导致查询速度较慢、查询效率较低。

[0055] 因此,本公开实施例提供了一种可以根据查询词确定目标字段类别,然后在目标字段类别信息中确定与查询词相关的查询信息的查询方法,该方法可以减少查询结果中的冗余信息,并进一步的提高查询效率。

[0056] 图3是根据一示例性实施例示出的一种查询方法的流程图。本公开实施例所提供的方法可以由任意具备计算处理能力的电子设备处理,例如上述图1实施例中的服务器105和/或终端设备102、103,在下面的实施例中,以服务器105为执行主体为例进行举例说明,但本公开并不限于此。

[0057] 参照图3,本公开实施例提供的查询方法可以包括以下步骤。

[0058] 在步骤S1中,获取查询词。

[0059] 在一些实施例中,查询词可以是目标用户输入的任意内容,可以是一个词语,可以是一个句子,也可以是一段话,本公开对此不做限制。例如,所述查询词可以是“糖尿病”、“高血压”等词语,也可以是“房颤心衰肺栓塞”、“右侧小腿损失”等短语,还可以是“急诊头部CT示考虑蛛网膜下腔出血,建议转入综合医院神经外科继续治疗”。

[0060] 在步骤S2中,通过目标神经网络模型对所述查询词进行处理,以确定与所述查询词匹配的目标字段类别。

[0061] 在一些实施例中,所述目标神经网络可以指的是fastText(一款集word2vec、文本分类等一体的机器学习训练工具,fastText模型具有分类精度高且训练模型速度非常快的优点),也可以指的是其它分类网络,例如、TextCNN(文本卷积网络)、循环神经网络模型等。

[0062] 在一些实施例中,医疗数据库中的电子病历可以包括多个字段,每个字段均可以包括自己的字段类别信息。

[0063] 例如,在一张电子病历中可以包括姓名字段(如“姓名:张三”)、既往史字段(如“既往史:否认高血压、冠心病、肝炎等”)、过敏史字段、诊断字段(如诊断:诊断为高血压2期)以及门诊医嘱字段等。

[0064] 可以理解的是,可以将字段中的字段名称作为该字段的字段类别,例如对于“既往史:否认高血压、冠心病、肝炎等”字段来说,“既往史”可以作为该字段的字段类别;对于“手术:全膝关节假体翻修术”字段来说,“手术”可以作为该字段的字段类别。

[0065] 在另外一些实施例中,还可以按照其它的方式对字段进行分类,本公开对此不做限制。

[0066] 在一些实施例中,所述目标字段类别可以指的是与所述查询词可能相关的字段类

别,本公开对此不做限制。

[0067] 在一些实施例中,所述目标字段类别可以由人为指定,也可以由目标神经网络模型根据查询词确定,还可以是服务器根据查询词进行类别匹配确定(即事先可以指定医疗词汇与医疗类别的匹配关系,服务器根据所述匹配关系确定所述查询词对应的目标字段类别)。

[0068] 在一些实施例中,目标用户可以在所述目标字段类别中根据查询词进行查询,以缩小查询范围,提供信息命中的准确率。

[0069] 在医疗领域中,可以事先将电子病历中的字段分为多个类别,例如可以将电子病历中的字段分为诊断、物理检查、实验室检查、手术、药品名称、症状以及科室等类别。

[0070] 在一些实施例中,所述目标字段类别可以是一个也可以是多个,本公开对此不做限制。

[0071] 例如,可以指定在上述诊断类类别的信息中进行信息的查询,可以在上述物理检查类别的信息中进行信息的查询,还可以同时在手术类别的信息和药品信息类别的中进行信息的查询。

[0072] 另外,所述目标字段类别还可以包括多个子字段类别,本公开对此不做限制。

[0073] 例如,当根据所述查询词确定的目标字段类别为诊断时,可以在入院诊断、出院诊断、病理诊断等多个医疗类别中进行信息的查询。

[0074] 在步骤S3中,根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。

[0075] 在医疗领域中,可以根据不同的医疗数据生成不同的数据库,例如可以根据电子病历生成电子病历数据库。

[0076] 在一些实施例中,所述目标医疗数据库可以指的是所述电子病历数据库。

[0077] 本实施例提供的查询方法,通过查询词确定了目标字段类别,并进一步的在目标字段类别中确定了与查询词相关的查询信息。该方法一方面相比于在包括所有医疗数据的数据库中进行信息查询的方法,可以极大地提高查询效率,减少不必要的资源浪费;另一方面可以精准定位至目标字段类别中进行信息的查询,以准确获取与查询词相关的查询信息,提供查询命中率。例如,当目标用户输入“高血压”时,根据本方案提供的查询方法可能会查到与治疗高血压相关的治疗信息,而不会是“高血压病房”等无关信息。

[0078] 图4是图3中步骤S2在一示例性实施例中的流程图。参考图4,上述步骤S2可以包括以下步骤。

[0079] 在步骤S21中,获取目标字段样本,所述目标字段样本包括字段类别信息。

[0080] 在一些实施例中,所述目标医疗数据库中的离线电子病历中可能会包括多个字段,所述电子病历中所有可能被检索的字段建立索引。

[0081] 在一些实施例中,上述离线的电子病历可以包括患者基本信息文本、患者院内基本信息文本、患者住院病案首页文本、患者临床版住院病案首页文本、患者就诊信息文本、患者转科历史文本、患者扩展文档文本、患者转床记录文本、患者科室就诊信息文本、患者诊断记录文本、医疗发票记录文本、医疗费用明细信息文本、患者中西医处方文本、患者药品类医嘱文本、患者的草药处方文本等不同类别的医疗文本。其中,每种类别的医疗文本样本又可以包括多个不同科室的医疗文本样本,每个科室的医疗文本样本又可以包括不同患

者的医疗文本样本。例如,在XX医院中,可能包括病历文本、住院文本、就论文本、转科文本等医疗文本,其中病历文本可能会包括传染科病历文本、呼吸内科病历文本、心血管内科病历文本等,而传染病科病历文本又可例如包括“张三病历文本”、“李四病历文本”等不同患者的病历文本。

[0082] 在一些实施例中,所述目标字段样本既可以包括所述电子病历的字段名称也可以包括所述电子病历的值。

[0083] 在电子病历中的“姓名:张三”、“性别:男”、“既往史:否认高血压、冠心病史,否认肝炎…”等都可以是一个字段。其中对于字段“姓名:张三”来说,“张三”就可以是该字段的内容;对于字段“既往史:否认高血压、冠心病史,否认肝炎…”来说“否认高血压、冠心病史,否认肝炎…”就可以是该字段的值。

[0084] 在一些实施例中,可以从上述电子病历中抽取所有可能被检索的字段作为目标字段样本,并可以对所述目标字段进行分类,例如可以按照目标字段的名称对目标字段进行分类,如分为诊断、药品、手术、物理检查、实验室检查等中的一类或者多类。

[0085] 在步骤S22中,根据所述目标字段样本以及所述目标字段样本的医疗类别信息训练所述目标神经网络模型。

[0086] 本实施例通过包括医疗类别信息的目标字段样本训练了目标神经网络模型,以便根据查询词确定目标字段类别。

[0087] 图5是图4中步骤S22在一示例性实施例中的流程图。参考图5,上述步骤S6可以包括以下步骤。

[0088] 在步骤S221中,根据N-gram方法对所述目标字段样本进行向量化处理,以获得目标向量。

[0089] 在一些实施例中,可以根据对包括类别信息的目标字段样本进行分词处理,如表1所示。

[0090] 表1

目标字段样本	分词后	类别标签
肝硬化腹水可能	肝硬化腹水可能	诊断
颅脑CT平扫颅脑	颅脑CT平扫颅脑	物理检查
乙肝表面抗体(定量)	乙肝表面抗体(定量)	实验室检查
右侧小腿损伤	右侧小腿损伤	诊断
全膝关节假体翻修术	全膝关节假体翻修术	手术
房颤心衰肺栓塞后	房颤心衰肺栓塞后	诊断
全血葡萄糖(餐后2小时)	全血葡萄糖(餐后2小时)	实验室检查
牛磺酸颗粒	牛磺酸颗粒	药品

[0092] 在一些实施例中,可以对所述目标字段样本进行向量化,以获得所述目标字段样本对应的目标向量。

[0093] 由于常用的词袋模型会丢失了词的顺序特征,所以这里还可以使用N-gram方法对所述目标字段样本进行向量化,以保留上下文词顺序的特征。

[0094] 在一些实施例中,可以将文本中的词的向量与根据N-gram方法获得的N-gram特征向量拼接后取平均值,以作为输入层的输入向量。

[0095] 例如,假设可以将所述目标字段样本分为 $w_1$ 、 $w_2$ 、 $w_3$ 等三个词N-gram向量可以表示为 $w_1w_2$ 、 $w_2w_3$ ,则根据N-gram方法生成的目标字段样本的目标向量可表示为公式(1)。

[0096]  $X=1/5(w_1+w_2+w_3+w_1w_2+w_1w_3)$  (1)

[0097] 由于N-gram向量远比目标字段样本的分词大的多,因此可以把所有的N-gram向量哈希到N个桶中,N为大于等于1的正整数。

[0098] 可以理解的是,映射到一个桶的N-gram向量共享一个向量。如图6所示,目标神经网络模型的最终输入向量大小为(分词数目b+哈希同数N)+向量维度。

[0099] 在步骤S222中,根据所述目标字段样本的目标向量以及所述目标字段样本的类别信息训练所述目标神经网络模型。

[0100] 如图6所示,可以将目标向量 $X_1 \sim X_n$ 输入至目标神经网络的输入向量层701,并进一步经过神经网络的隐藏层702并经过分类层703,以获得预测的分类结果。可以理解的是,根据目标向量预测的分类结果以及目标字段类别可以更新所述神经网络的参数,以完成所述神经网络的训练。

[0101] 本实施例提供的查询方法,基于N-gram方法确定的目标向量,充分考虑了上下文之间的关系,使得根据目标向量训练的目标神经网络模型在根据查询词进行类别确认时可以更加准确。

[0102] 图7是图3中步骤S3在一示例性实施例中的流程图。参考图7,上述步骤S3可以包括以下步骤。

[0103] 在步骤S31中,确定所述查询词的相关词。

[0104] 在一些实施例中,所述查询词的相关词可以指的是所述查询词的同义词,也可以指的是所述查询词的子类词。其中,所述查询词的子类词可以指的是所述查询词的子类别。例如,对于查询词“糖尿病”来说,“糖尿病1期”、“糖尿病2期”、“特殊类型的糖尿病”、“妊娠期糖尿病”等都是“糖尿病”的子类词。

[0105] 在步骤S32中,根据所述查询词以及所述查询词的相关词在目标字段类别中确定与所述查询词匹配的信息。

[0106] 在一些实施例中,可以根据所述查询词、所述查询词的同义词以及所述查询词的子类词在目标字段类别中确定与所述查询词匹配的信息。

[0107] 在一些实施例中,不仅可以获得查询词在在目标字段类别中确定的匹配信息,还可以获得所述查询词的相关词(例如同义词或子类词)在目标字段类别中确定的匹配信息。

[0108] 本实施例提供的查询方法,不仅根据查询词确定查询结果,还根据查询词的同义词和子类别词进行查询结果的确定,适当的扩大了查询范围,提高了查询准确率。

[0109] 图8是图7中步骤S32在一示例性实施例中的流程图。参考图8,上述步骤S32可以包括以下步骤。

[0110] 在步骤S321中,对所述查询词进行分词处理,以获得所述查询词的分词结果。

[0111] 在一些实施例中,可以对查询词进行分词处理,例如对查询词“肺癌”进行分词可以获得“肺”和“癌”两个词。

[0112] 在步骤S322中,对所述查询词的相关词进行分词处理,以获得所述相关词的分词结果。

[0113] 假设所述查询词为“肺癌”,所述查询词的同义词为“肺恶性肿瘤”以及子类词例如

“左肺恶性肿瘤”、“右肺恶性肿瘤”、“双肺恶性肿瘤”以及“肺多处恶性肿瘤”等。

[0114] 对所述查询词的相关词(例如同义词和子类词)进行分词处理可以获得如表2所示的分词结果。

[0115] 表2

查询词的相关词	分词结果
肺恶性肿瘤	肺恶性肿瘤
左肺恶性肿瘤	左肺恶性肿瘤
右肺恶性肿瘤	右肺恶性肿瘤
双肺恶性肿瘤	双肺恶性肿瘤
肺多处恶性肿瘤	肺多处恶性肿瘤

[0117] 在步骤S323中,根据所述查询词的分词结果和所述相关词的分词结果在所述目标字段类别中确定与所述查询词匹配的信息。

[0118] 本实施例提供的查询方法,对查询词及其相关词进一步的进行了分词处理,并根据分词结果进行查询,进一步的扩大了查询范围、提高了查询准确率。

[0119] 在另外一些实施例中,所述目标字段类别包括第一字段类别和第二字段类别。图3中步骤S3还可以包括以下步骤:根据所述第一字段类别在所述目标医疗数据库中确定与所述查询词匹配的第一结果;根据所述第二字段类别在所述目标医疗数据库中确定与所述查询词匹配的第二结果;对所述第一结果和所述第二结果进行去重处理,以确定与所述查询词匹配的查询结果。

[0120] 例如,假设所述目标字段类别为诊断类别,那么所述诊断类别还可以进一步的包括入院诊断类别、出院诊断类别、检测诊断类别、病理诊断类别等;再例如,所述目标字段类别可以包括诊断类别、医嘱类别、治疗类别以及药品类别等。

[0121] 在一些实施例中,可以根据至少个字段类别在所述目标医疗数据库中确定与所述查询词相关的匹配结果。

[0122] 可以理解的是,不根据不同的字段类别进行查询能命中同一病历,所以在根据不同的字段类别进行查询之后,还需要对查询结果进行去重处理,以确定最后的命中的电子病历。

[0123] 图9是根据本公开实施例示出的一种查询方法的流程图。参考图10,本公开实施例提供的查询方法可以包括以下步骤。

[0124] 在步骤S101中,获取查询词。

[0125] 在步骤S102中,通过目标神经网络模型对查询词进行处理以确定目标字段类别。

[0126] 在步骤S103中,获取所述查询词的同义词。

[0127] 在步骤S104中,获取所述查询词的子类词。

[0128] 在步骤S105中,对所述查询词、所述查询词的同义词和子类词进行分词处理。

[0129] 在步骤S106中,根据所述查询词、查询词的同义词和子类词的分词结果在所述目标字段类别中进行查询,以获得查询结果。

[0130] 在一些实施例中,所述查询结果可以是目标医疗数据库中与所述查询词匹配的字段。

[0131] 在步骤S107中,根据查询结果进行倒排索引,以确定与所述查询词相关的电子病

历。

[0132] 在一些实施例中,目标医疗数据库中的医疗字段与电子病历存在对应的关系。当确定了目标医疗数据库中的医疗字段时,那么对应的电子病历也可以被命中。

[0133] 本实施例提供的查询方法,通过查询词确定了目标字段类别,并进一步的在目标字段类别中确定了与查询词相关的查询信息。该方法一方面相比于在包括所有医疗类别数据的数据库中进行信息查询的方法,可以极大地提高查询效率,减少不必要的资源浪费;另一方面可以精准定位至目标字段类别中进行信息的查询,以精准获取与查询词相关的查询信息。例如,当目标用户输入“高血压”时,根据本方案提供的查询方法可能会查到与治疗高血压相关的治疗信息,而不会是“高血压病房”等无关信息。

[0134] 图10是根据一示例性实施例示出的一种查询装置的框图。参照图10,本公开实施例提供的查询装置1100可以查询词获取模块1101、目标字段类别确定模块1102以及查询结果获取模块1103。

[0135] 其中,所述查询词获取模块1101可以配置为获取查询词;所述目标字段类别确定模块1102可以配置为通过目标神经网络模型对所述查询词进行处理,以确定与所述查询词匹配的目标字段类别;所述查询结果获取模块1103可以配置为根据所述目标字段类别在目标医疗数据库中确定与所述查询词匹配的查询结果。

[0136] 在一些实施例中,所述查询装置1100还可以包括:医疗样本确定模块、训练模块。其中,所述医疗样本确定模块可以配置为获取目标字段样本,所述目标字段样本包括医疗类别信息,所述目标字段样本取自于医疗病历;所述训练模块可以配置为根据所述目标字段样本以及所述目标字段样本的医疗类别信息训练所述目标神经网络模型。

[0137] 在一些实施例中,所述训练模块可以包括向量生成单元、模型训练单元。

[0138] 其中,所述向量生成单元可以配置为根据N-gram方法生成所述目标字段样本的向量;所述模型训练单元可以配置为根据所述目标字段样本的向量以及所述目标字段样本的类别信息训练所述目标神经网络模型。

[0139] 在一些实施例中,所述查询结果获取模块1103可以包括相关词确定单元和匹配信息确定单元。

[0140] 其中,所述相关词确定单元可以配置为确定所述查询词的相关词;所述匹配信息确定单元可以配置为根据所述查询词以及所述查询词的相关词在目标字段类别中确定与所述查询词匹配的信息。

[0141] 在一些实施例中,所述查询词的相关词包括所述查询词的同义词以及所述查询词的子类词。

[0142] 在一些实施例中,所述匹配信息确定单元可以配置为根据所述查询词、所述查询词的同义词以及所述查询词的子类词在目标字段类别中确定与所述查询词匹配的信息。

[0143] 在一些实施例中,所述匹配信息确定单元可以包括第一分词子单元、第二分词子单元以及匹配信息获取子单元。

[0144] 其中,所述第一分词单元可以配置为对所述查询词进行分词处理,以获得所述查询词的分词结果;所述第二分词单元可以配置为对所述查询词的相关词进行分词处理,以获得所述相关词的分词结果;所述匹配信息获取子单元可以配置为根据所述查询词的分词结果和所述相关词的分词结果在所述目标字段类别中确定与所述查询词匹配的信息。

[0145] 由于本公开的示例实施例的查询装置1100的各个功能模块与上述查询方法的示例实施例的步骤对应,因此在此不再赘述。

[0146] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,本公开实施例的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中,包括若干指令用以使得一台计算设备(可以是个人计算机、服务器、移动终端、或者智能设备等)执行根据本公开实施例的方法,例如图3的一个或多个所示的步骤。

[0147] 此外,上述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,上述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

[0148] 本领域技术人员在考虑说明书及实践这里公开的公开后,将容易想到本公开的其他实施例。本公开旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未申请的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由权利要求指出。

[0149] 应当理解的是,本公开并不限于这里已经示出的详细结构、附图方式或实现方法,相反,本公开意图涵盖包含在所附权利要求的精神和范围内的各种修改和等效设置。

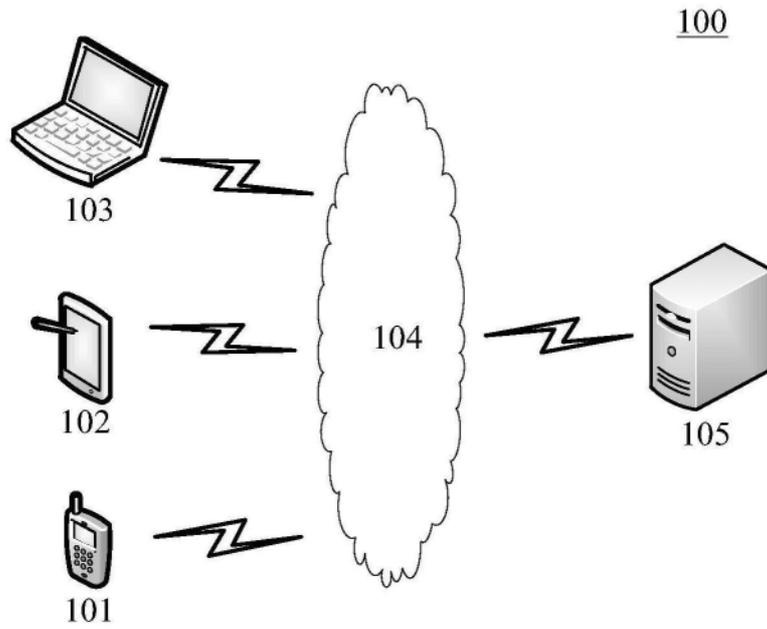


图1

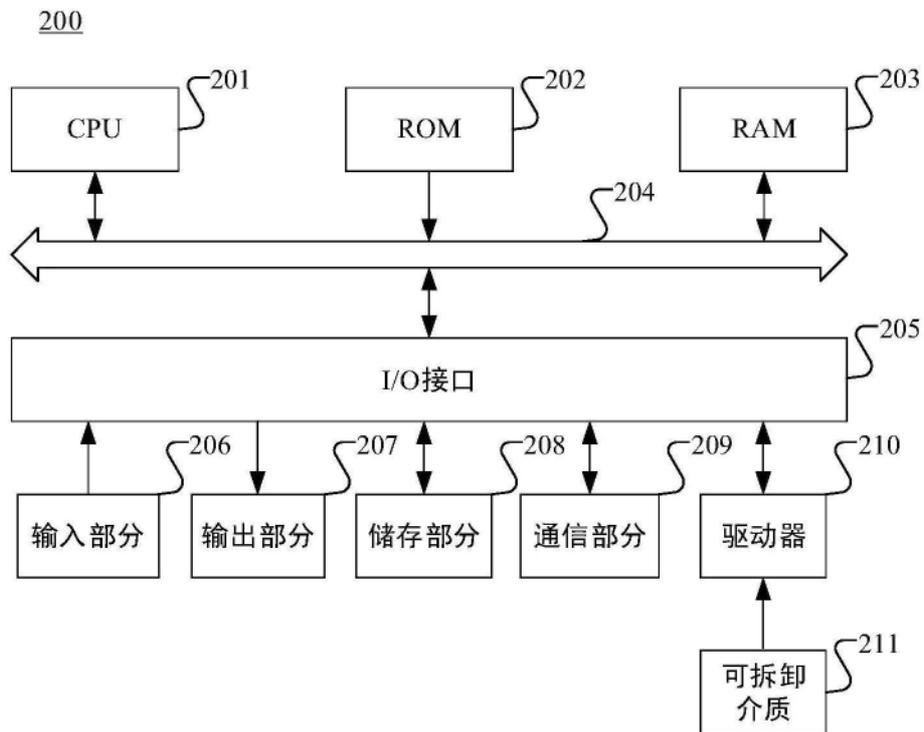


图2

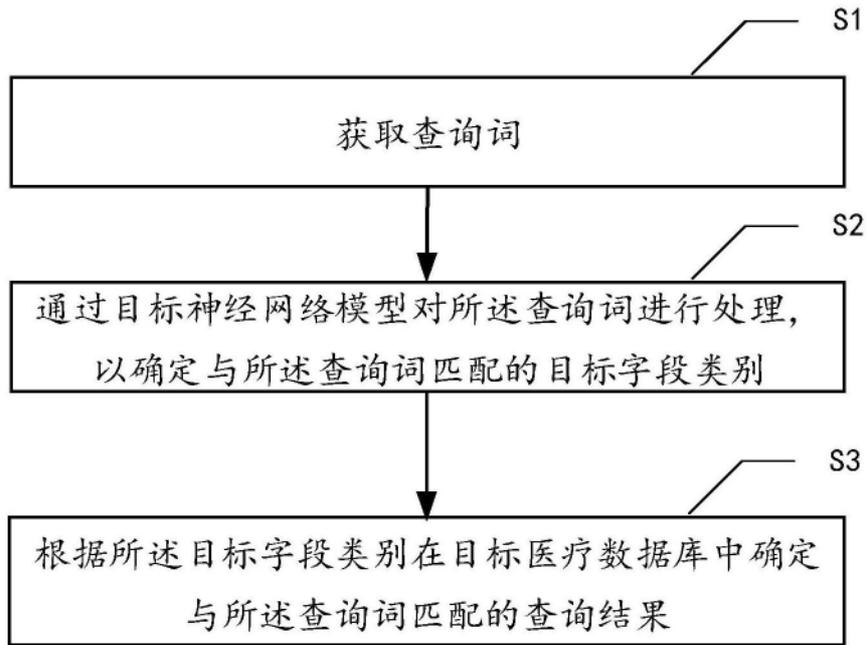


图3

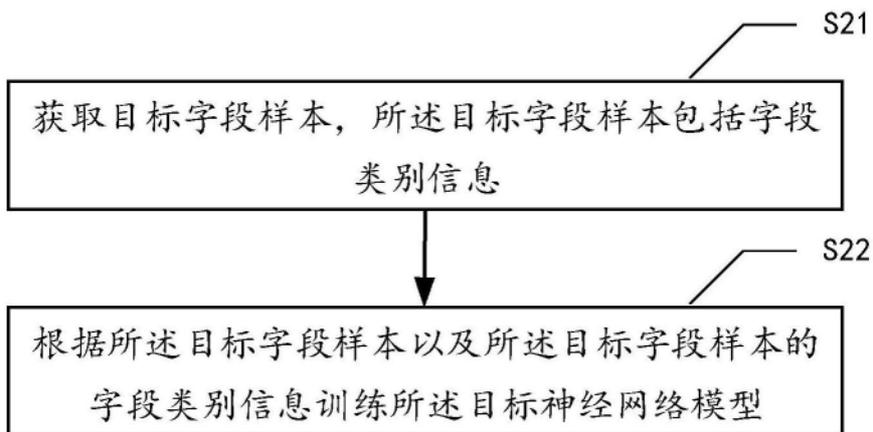


图4

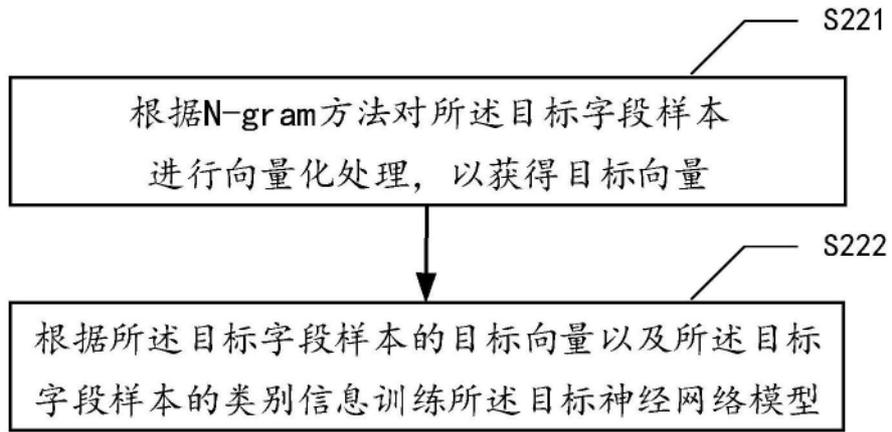


图5

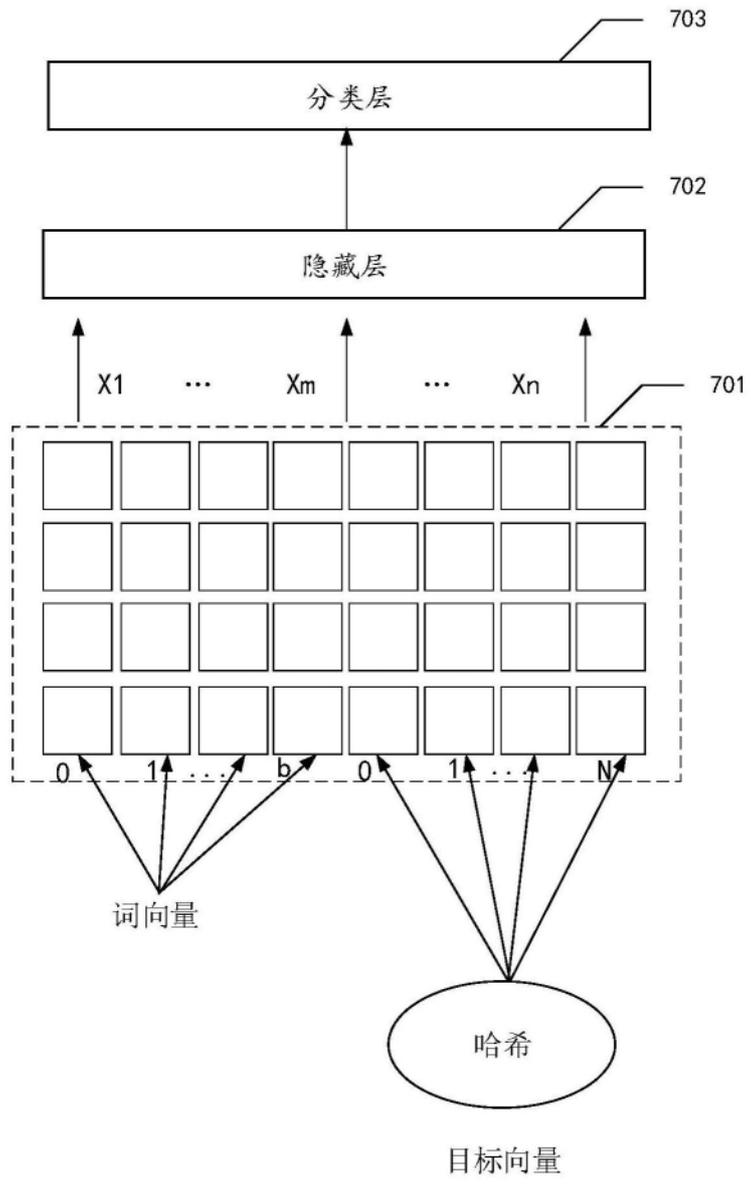


图6

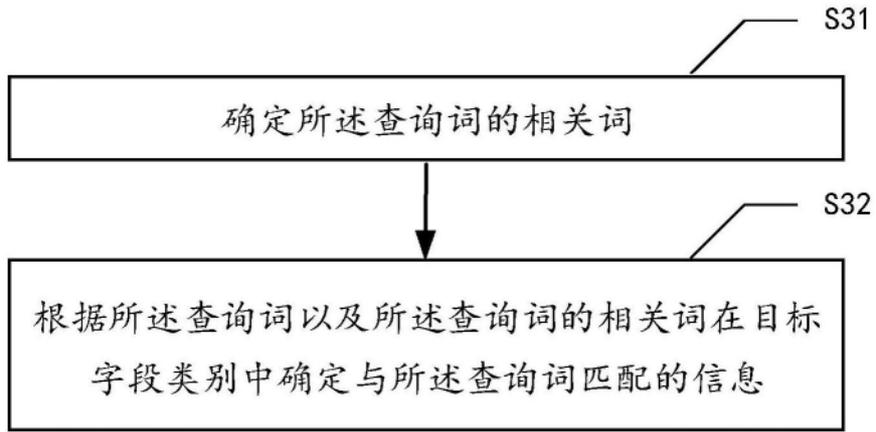


图7

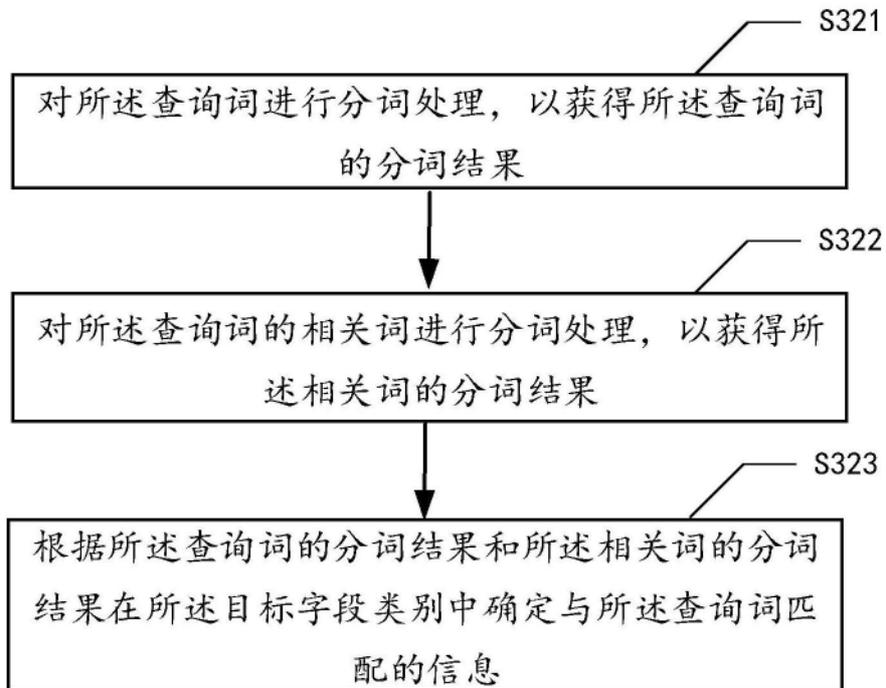


图8

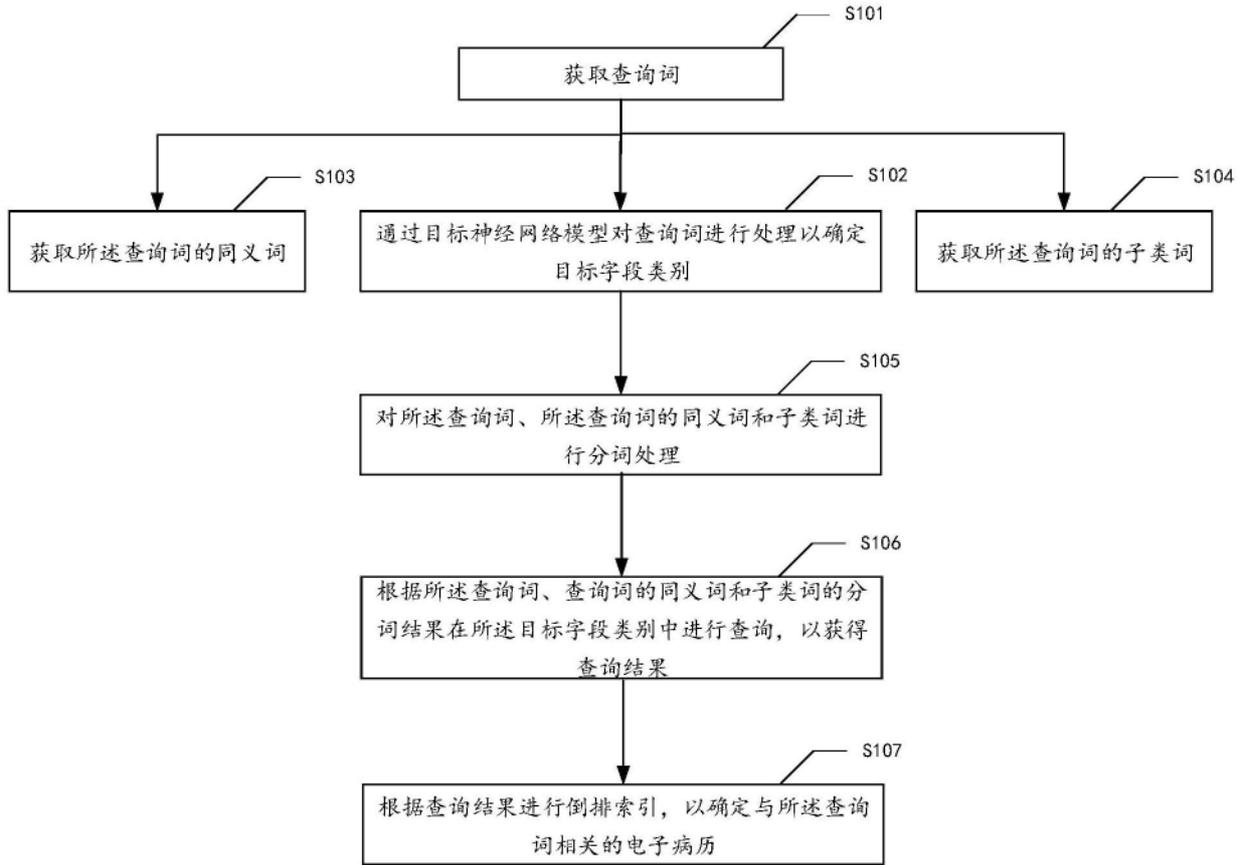


图9

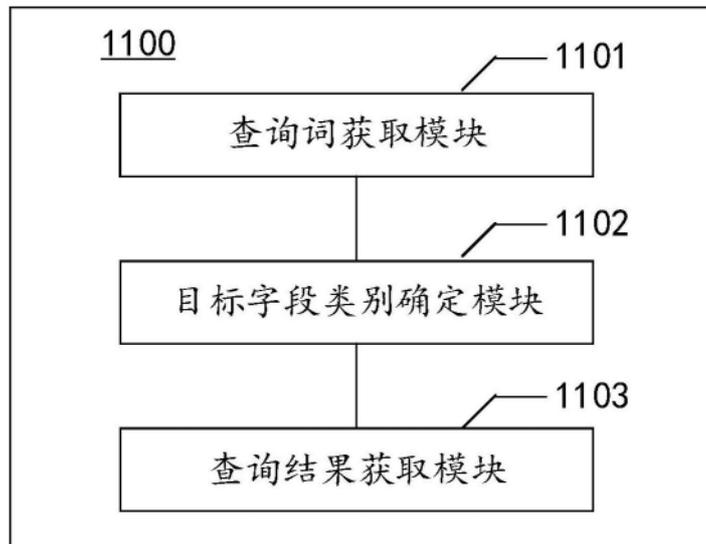


图10