



(12) 发明专利申请

(10) 申请公布号 CN 119740576 A

(43) 申请公布日 2025. 04. 01

(21) 申请号 202411635595.8

(22) 申请日 2024.11.15

(71) 申请人 河南中医药大学

地址 450000 河南省郑州市郑东新区金水
东路156号

(72) 发明人 任妍 黄静 段涛 付玉 习书晗

(74) 专利代理机构 河南豫龙律师事务所 41177
专利代理师 张电臣

(51) Int. Cl.

G06F 40/289 (2020.01)

G06F 40/284 (2020.01)

G06N 5/025 (2023.01)

G06F 18/22 (2023.01)

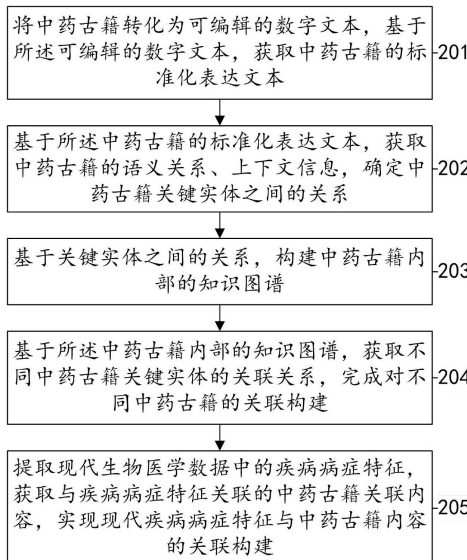
权利要求书3页 说明书10页 附图7页

(54) 发明名称

一种中药古籍的关联构建方法、装置、设备
及存储介质

(57) 摘要

本申请提供了一种中药古籍的关联构建方法、装置、设备及存储介质,所述方法包括:将中药古籍转化为可编辑的数字文本,获取中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系,构建中药古籍内部的知识图谱;基于中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,提取现代生物学数据中的疾病病症特征,匹配相对应的中药古籍内容,实现疾病病症特征与中药古籍内容的关联构建,本申请通过不同中药古籍之间的关联构建,实现对中药知识的深入分析,并将中药古籍中的内容与现代生物学数据中的病症特征进行关联匹配,使得中药古籍重新发挥其重要作用。



1. 一种中药古籍的关联构建方法,其特征在于,包括:

将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本;

基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系;

基于关键实体之间的关系,构建中药古籍内部的知识图谱;

基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建;

提取现代生物学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建。

2. 根据权利要求1所述的中药古籍的关联构建方法,其特征在于,所述将中药古籍转化为可编辑的数字文本,具体内容包括:

获取不同的中药古籍,包括现有的中药古籍文献资料;

将所述不同的中药古籍进行扫描,生成高清电子图像文件集;

基于光学字符识别技术将所述电子图像文件集转化为可编辑的数字文本集;

将所述可编辑的数字化文本进行格式化处理,消除识别错误,并转化为统一的预设格式文本。

3. 根据权利要求2所述的中药古籍的关联构建方法,其特征在于,所述基于所述可编辑的数字文本,获取中药古籍的标准化表达文本,具体内容包括:

构建中医药术语映射知识库,将不同表达的术语映射为统一标准术语;

对药材名称、疾病名称进行规范化处理,建立药材-疾病-方剂的标准对应关系;

识别所述可编辑的数字文本中药材、疾病、方剂关键实体,对识别的关键实体进行语义类型标注;

基于中医药术语映射知识库,将识别出的关键实体替换为统一标准术语,将替换后的统一标准术语进行标注,生成标准化的文本表达。

4. 根据权利要求1所述的中药古籍的关联构建方法,其特征在于,所述基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系,具体内容包括:

基于中药古籍的语言特点和专业知识,制定语义关系抽取规则;

基于正则表达式及语义关系抽取规则,通过定义包含关系关键词和实体类型的正则表达式模式,获取关键实体之间的语义关系;

对于每个关键实体,设定一个固定的文本窗口大小,在标准化表达文本中,提取关键实体周围窗口内的词汇作为上下文信息;

将获取的语义关系信息和上下文信息进行整合,使用数据结构对整合后的语义关系和上下文信息进行存储。

5. 根据权利要求1所述的中药古籍的关联构建方法,其特征在于,所述基于关键实体之间的关系,构建中药古籍内部的知识图谱,具体内容包括:

分析关键实体的种类,确定知识图谱中的节点类型;

根据关键实体之间确定的关系,定义知识图谱中边的类型;

基于图数据库的操作语句或对应的API及确定好的节点类型,逐个创建知识图谱中的节点;

基于确定好的边的类型,在相应的节点之间创建边来表示关键实体之间的关系;

在创建节点和边的过程中,同时设置好它们预先规划的属性值,除在创建节点和边时直接设置部分属性外,还可以通过后续更新操作来完善节点和边的属性设置。

6.根据权利要求1所述的中药古籍的关联构建方法,其特征在于,所述基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,具体内容包括:

识别出不同中药古籍内部知识图谱的关键实体;

获取已识别关键实体之间的语义关系;

基于识别的关键实体以及关键实体之间的语义关系,将每本中药古籍视为关联网络中的一个节点,把不同古籍知识图谱中的关键实体也作为节点,通过对关键实体添加古籍来源属性区分不同古籍中的同名实体,确保网络中节点的唯一性和可识别性;依据语义关系定义边,当两个关键实体来自不同古籍且存在语义关系,则在对应的古籍节点以及关键实体节点之间建立边,以此构建不同中药古籍之间的关联网络。

7.根据权利要求1所述的中药古籍的关联构建方法,其特征在于,所述提取现代生物学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建,具体表现为:

基于现代生物学数据,采集与疾病病症相关的数据;

将所述与疾病病症相关的数据作为预设疾病特征模型的输入,获取疾病病症特征;

基于预设术语映射表,获取疾病病症特征与中药古籍中的相关表述,将所述相关表述作为检索条件,获取与所述疾病病症特征相关的中药古籍内容节点;

提取中药古籍内容节点的相关信息,作为与疾病病症特征关联的中药古籍关联内容;

利用词向量模型将疾病病症特征描述和中药古籍内容分别转化为词向量序列,然后计算两者之间的余弦相似度语义相似度指标,根据设定的相似度阈值,判断语义是否匹配;

当疾病病症特征与中药古籍内容满足语义匹配后,将现代医学中的症状进行编码处理,将古籍中的症状描述通过自然语言处理技术转化为词向量,通过计算疾病病症特征与中药古籍内容之间的相似性指标,判断症状的相似性程度;

当疾病病症特征与中药古籍内容症状的相似性程度满足预设阈值时,将疾病病症特征与中药古籍内容在知识图谱层面上实现关联构建,形成一个跨领域的知识网络。

8.一种中药古籍的关联构建装置,用于实现权利要求1-7的中药古籍的关联构建方法,其特征在于,包括:

标准化表达文本获取模块,用于将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本;

关键实体关系确定模块,用于基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系;

知识图谱构建模块,用于基于关键实体之间的关系,构建中药古籍内部的知识图谱;

第一关联构建模块,用于基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建;

关联构建模块模块,用于提取现代生物医学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建。

9.一种电子设备,其特征在于,包括:

一个或多个处理器;存储器;以及一个或多个计算机程序,其中所述一个或多个计算机程序被存储在所述存储器中,所述一个或多个计算机程序包括指令,当所述指令被所述设备执行时,使得所述设备执行如权利要求1至7任一项所述的中药古籍的关联构建方法的步骤。

10.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有计算机程序,当其在计算机上运行时,使得计算机执行如权利要求1至7任一项所述的中药古籍的关联构建方法的步骤。

一种中药古籍的关联构建方法、装置、设备及存储介质

技术领域

[0001] 本申请涉及中药古籍关联构建技术领域,具体说是一种中药古籍的关联构建方法、装置、设备及存储介质。

背景技术

[0002] 目前对中药古籍进行数字化研究比较普遍,单纯的数字化存储虽然便于保存和检索,但无法直观展现古籍间的内在联系和深层逻辑,这种局限性使得古籍知识的有效整合和高效利用变得尤为困难,并且古籍采用文言文撰写,术语复杂且历经多个历史阶段的演变,这使得不同古籍间的表达方式、理论架构和药物认知产生了显著差异。这种多样性不仅增加了理解的难度,也阻碍了知识的连贯性,古籍间的联系往往微妙且复杂,人工识别这些关联不仅耗时费力,还要求研究者具备深厚的中药学知识和丰富的实践经验,进一步加剧了古籍知识整合的难度。因此如何对中药古籍进行关联构建便成为需要重点解决的问题。

发明内容

[0003] 为了克服现有技术中存在的上述问题,本申请提供了一种中药古籍的关联构建方法、装置、设备及存储介质,采用了如下的技术方案:

[0004] 第一方面,本申请提供了一种中药古籍的关联构建方法,包括:

[0005] 将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本;

[0006] 基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系;

[0007] 基于关键实体之间的关系,构建中药古籍内部的知识图谱;

[0008] 基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建;

[0009] 提取现代生物医学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建。

[0010] 进一步地,所述将中药古籍转化为可编辑的数字文本,具体内容包括:

[0011] 获取不同的中药古籍,包括现有的中药古籍文献资料;

[0012] 将所述不同的中药古籍进行扫描,生成高清电子图像文件集;

[0013] 基于光学字符识别技术将所述电子图像文件集转化为可编辑的数字文本集;

[0014] 将所述可编辑的数字化文本进行格式化处理,消除识别错误,并转化为统一的预设格式文本。

[0015] 进一步地,所述基于所述可编辑的数字文本,获取中药古籍的标准化表达文本,具体内容包括:

[0016] 构建中医药术语映射知识库,将不同表达的术语映射为统一标准术语;

[0017] 对药材名称、疾病名称进行规范化处理,建立药材-疾病-方剂的标准对应关系;

[0018] 识别所述可编辑的数字文本中药材、疾病、方剂关键实体,对识别的关键实体进行语义类型标注;

[0019] 基于中医药术语映射知识库,将识别出的关键实体替换为统一标准术语,将替换后的统一标准术语进行标注,生成标准化的文本表达。

[0020] 进一步地,所述基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系,具体内容包括:

[0021] 基于中药古籍的语言特点和专业知识,制定语义关系抽取规则;

[0022] 基于正则表达式及语义关系抽取规则,通过定义包含关系关键词和实体类型的正则表达式模式,获取关键实体之间的语义关系;

[0023] 对于每个关键实体,设定一个固定的文本窗口大小,在标准化表达文本中,提取关键实体周围窗口内的词汇作为上下文信息;

[0024] 将获取的语义关系信息和上下文信息进行整合,使用数据结构对整合后的语义关系和上下文信息进行存储。

[0025] 进一步地,所述基于关键实体之间的关系,构建中药古籍内部的知识图谱,具体内容包括:

[0026] 分析关键实体的种类,确定知识图谱中的节点类型;

[0027] 根据关键实体之间确定的关系,定义知识图谱中边的类型;

[0028] 基于图数据库的操作语句或对应的API及确定好的节点类型,逐个创建知识图谱中的节点;

[0029] 基于确定好的边的类型,在相应的节点之间创建边来表示关键实体之间的关系;

[0030] 在创建节点和边的过程中,同时设置好它们预先规划的属性值,除在创建节点和边时直接设置部分属性外,还可以通过后续更新操作来完善节点和边的属性设置。

[0031] 进一步地,所述基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,具体内容包括:

[0032] 识别出不同中药古籍内部知识图谱的关键实体;

[0033] 获取已识别关键实体之间的语义关系;

[0034] 基于识别的关键实体以及关键实体之间的语义关系,将每本中药古籍视为关联网络中的一个节点,把不同古籍知识图谱中的关键实体也作为节点,通过对关键实体添加古籍来源属性区分不同古籍中的同名实体,确保网络中节点的唯一性和可识别性;依据语义关系定义边,当两个关键实体来自不同古籍且存在语义关系,则在对应的古籍节点以及关键实体节点之间建立边,以此构建不同中药古籍之间的关联网络。

[0035] 进一步地,所述基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,还包括:

[0036] 基于构建的关联网络和已有的知识图谱,从一个关键实体节点出发,通过深度优先搜索与关键节点相关的其他关键实体节点,分析关键节点与其他关键实体节点之间的关系和路径,获取关键节点与其他关键实体节点潜在的关联关系。

[0037] 进一步地,所述提取现代生物医学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建,具体表现为:

[0038] 基于现代生物医学数据,采集与疾病病症相关的数据;

[0039] 将所述与疾病病症相关的数据作为预设疾病特征模型的输入,获取疾病病症特征;

[0040] 基于预设术语映射表,获取疾病病症特征与中药古籍中的相关表述,将所述相关表述作为检索条件,获取与所述疾病病症特征相关的中药古籍内容节点;

[0041] 提取中药古籍内容节点的相关信息,作为与疾病病症特征关联的中药古籍关联内容;

[0042] 利用词向量模型将疾病病症特征描述和中药古籍内容分别转化为词向量序列,然后计算两者之间的余弦相似度语义相似度指标,根据设定的相似度阈值,判断语义是否匹配;

[0043] 当疾病病症特征与中药古籍内容满足语义匹配后,将现代医学中的症状进行编码处理,将古籍中的症状描述通过自然语言处理技术转化为词向量,通过计算疾病病症特征与中药古籍内容之间的相似性指标,判断症状的相似性程度;

[0044] 当疾病病症特征与中药古籍内容症状的相似性程度满足预设阈值时,将疾病病症特征与中药古籍内容在知识图谱层面上实现关联构建,形成一个跨领域的知识网络。

[0045] 方面,本申请还提供了一种中药古籍的关联构建装置,包括:

[0046] 标准化表达文本获取模块,用于将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本;

[0047] 关键实体关系确定模块,用于基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系;

[0048] 知识图谱构建模块,用于基于关键实体之间的关系,构建中药古籍内部的知识图谱;

[0049] 第一关联构建模块,用于基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建;

[0050] 关联构建模块模块,用于提取现代生物医学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建。

[0051] 第三方面,本申请提供了一种电子设备,包括:

[0052] 一个或多个处理器;存储器;以及一个或多个计算机程序,其中所述一个或多个计算机程序被存储在所述存储器中,所述一个或多个计算机程序包括指令,当所述指令被所述设备执行时,使得所述设备执行如第一方面所述的方法。

[0053] 第四方面,本申请提供了一种计算机可读存储介质,该计算机可读存储介质中存储有计算机程序,当其在计算机上运行时,使得计算机执行如第一方面所述的方法。

[0054] 第五方面,本申请提供了一种计算机程序,当所述计算机程序被计算机执行时,用于执行第一方面所述的方法。

[0055] 在一种可能的设计中,第五方面中的程序可以全部或者部分存储在与处理器封装在一起的存储介质上,也可以部分或者全部存储在不与处理器封装在一起的存储器上。

[0056] 本申请具有以下有益效果:

[0057] 1.本申请通过将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本,在进行不同中药古籍关联构建的过程中,可以采用标准化话的表达方式,打破不同中药古籍之间的差异性。

[0058] 2.本申请通过基于中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定关键实体之间的关系,基于关键实体之间的关系,构建中药古籍内部的知识图谱,基于中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,提取现代生物医学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建;本申请从标准化表达文本中获取语义关系和上下文信息,能够挖掘出古籍中隐藏的知识,在不同中药古籍之间建立关联关系,可以发现不同古籍在知识传承和发展过程中的联系,将现代生物医学数据与中药古籍内容关联构建,能够挖掘出传统中医药知识与现代医学知识的结合点;本申请通过不同中药古籍之间的关联构建,实现对中药知识的深入分析,并将中药古籍中的内容与现代生物医学数据中的病症特征进行关联匹配,使得中药古籍重新发挥其重要作用。

附图说明

- [0059] 图1为本申请实施例可以应用于其中的示例性系统架构图;
- [0060] 图2为本申请实施例的中药古籍的关联构建方法流程图;
- [0061] 图3为本申请实施例的获取可编辑数字化文本流程图;
- [0062] 图4为本申请实施例的标准化表达文本获取流程图;
- [0063] 图5为本申请实施例的中药古籍关键实体关系确定流程图;
- [0064] 图6为本申请实施例的中药古籍内部知识图谱构建流程图;
- [0065] 图7为本申请实施例的不同中药古籍关联构建流程图;
- [0066] 图8为本申请实施例的疾病病症特征与中药古籍内容的关联构建流程图;
- [0067] 图9为本申请实施例的装置流程图;
- [0068] 图10是本申请实施例的计算机设备示意图。

具体实施方式

[0069] 除非另有定义,本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同;本文中在申请的说明书中所使用的术语只是为了描述具体的实施例的目的,不是旨在于限制本申请;本申请的说明书和权利要求书及上述附图说明中的术语“包括”和“具有”以及它们的任何变形,意图在于覆盖不排他的包含。本申请的说明书和权利要求书或上述附图中的术语“第一”、“”等是用于区别不同对象,而不是用于描述特定顺序。

[0070] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0071] 为了使本技术领域的人员更好地理解本申请方案,下面将结合附图,对本申请实施例中的技术方案进行清楚、完整地描述。

[0072] 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。

[0073] 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。终端设备101、102、103上可以安装有各种通讯客户端应用,例如网页浏览器应用、购物类应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0074] 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、电子书阅读器、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。

[0075] 服务器105可以是提供各种服务的服务器,例如对终端设备101、102、103上显示的页面提供支持的后台服务器。

[0076] 需要说明的是,本申请实施例所提供的中药古籍的关联构建方法一般由服务器/终端设备执行,相应地,中药古籍的关联构建装置一般设置于服务器/终端设备中。

[0077] 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。根据实现需要,可以具有任意数目的终端设备、网络和服务器的。

[0078] 继续参考图2,图中示出了本申请的一种中药古籍的关联构建方法流程图,所述方法包括以下步骤:

[0079] 步骤201,将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本。

[0080] 在一种可能的实施方式中,所述将中药古籍转化为可编辑的数字化文本,请参考图3,具体表现为:

[0081] 步骤31,获取不同的中药古籍,包括现有的中药古籍文献资料。

[0082] 步骤32,将所述不同的中药古籍进行扫描,生成高清电子图像文件集。

[0083] 步骤33,基于光学字符识别技术将所述电子图像文件集转化为可编辑的数字文本集。

[0084] 步骤34,将所述可编辑的数字化文本进行格式化处理,消除识别错误,并转化为统一的预设格式文本。

[0085] 在一种可能的实施方式中,所述基于所述可编辑的数字文本,获取中药古籍的标准化表达文本,请参考图4,具体表现为:

[0086] 步骤41,构建中医药术语映射知识库,将不同表达的术语映射为统一标准术语。

[0087] 步骤42,对药材名称、疾病名称进行规范化处理,建立药材-疾病-方剂的标准对应关系。

[0088] 步骤43,识别所述可编辑的数字文本中药材、疾病、方剂关键实体,对识别的关键实体进行语义类型标注。

[0089] 步骤44,基于中医药术语映射知识库,将识别出的关键实体替换为统一标准术语,将替换后的统一标准术语进行标注,生成标准化的文本表达。

[0090] 步骤202,基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系。

[0091] 在一种可能的实施方式中,所述基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系,请参考图5,具体表现

为:

[0092] 步骤51,基于中药古籍的语言特点和专业知识,制定语义关系抽取规则;例如,对于治疗关系,如果文本中出现“治”“疗”“愈”“主治”等词汇,且前后分别为药材名称和病症名称,则可以判定存在治疗关系。如“人参治气虚”,就可以确定人参和气虚之间的治疗关系。

[0093] 步骤52,基于正则表达式及语义关系抽取规则,通过定义包含关系关键词和实体类型的正则表达式模式,获取关键实体之间的语义关系。例如,对于方剂和药材之间的包含关系,正则表达式可以设定为“方剂名[包含|有|由]药材名”。

[0094] 步骤53,对于每个关键实体,设定一个固定的文本窗口大小,在标准化表达文本中,提取关键实体周围窗口内的词汇作为上下文信息。例如,对于关键实体“人参”,如果文本是“人参味甘,性微温,大补元气”,以“人参”为中心,前后各取一定数量的词汇,得到的上下文信息就是“味甘,性微温,大补元气”。

[0095] 步骤54,将获取的语义关系信息和上下文信息进行整合,使用数据结构对整合后的语义关系和上下文信息进行存储。例如,对于两个关键实体(药材和病症)之间的治疗关系,结合上下文信息(如药材的性味、病症的症状等)可以更准确地理解这种治疗关系的具体情况。

[0096] 其中中药古籍关键实体之间的关系类型,包括药材-疾病关系、方剂-症状关系、药材-功效关系、疾病-治疗关系等。

[0097] 步骤203,基于关键实体之间的关系,构建中药古籍内部的知识图谱。

[0098] 在一种可能的实施方式中,所述基于关键实体之间的关系,构建中药古籍内部的知识图谱,请参考图6,具体内容包括:

[0099] 步骤61,分析关键实体的种类,确定知识图谱中的节点类型。在中药古籍的情境下,常见的节点类型可能包括:药材节点:代表各种中药材,其属性可以包含药材名称、性味、归经、产地、功效等信息。病症节点:对应古籍中提到的各类病症,属性可涵盖病症名称、症状表现、病因、分类等方面。方剂节点:用于表示古籍记载的方剂,属性有方剂名称、组成药材、用法用量、主治病症等。人物节点:若古籍涉及作者、医者等人物相关内容,可设立人物节点,属性如姓名、朝代、主要贡献等。炮制方法节点:针对中药的炮制相关描述,设立此节点,属性包括炮制方法名称、操作步骤、适用药材等。

[0100] 步骤62,根据关键实体之间确定的关系,定义知识图谱中边的类型。例如:治疗关系边:连接药材节点和病症节点,表示该药材可用于治疗对应的病症。组成关系边:存在于方剂节点和药材节点之间,表明该药材是组成此方剂的成分之一。撰写关系边:连接人物节点(作者)和古籍相关节点(如方剂节点、病症节点等,如果是该作者所撰写涉及的内容),体现创作关系。炮制关系边:关联药材节点和炮制方法节点,说明该药材可采用此炮制方法进行加工。

[0101] 步骤63,基于图数据库的操作语句或对应的API及确定好的节点类型,逐个创建知识图谱中的节点。

[0102] 步骤64,基于确定好的边的类型,在相应的节点之间创建边来表示关键实体之间的关系。例如,创建表示人参可治疗气虚的治疗关系边等。

[0103] 在一种可能的实施方式中,在创建节点和边的过程中,同时设置好节点和边预先

规划的属性值,除在创建节点和边时直接设置部分属性外,还可以通过后续更新操作来完善节点和边的属性设置。

[0104] 举例地,对于治疗关系边,可设置属性如“治疗依据”,记录在古籍中提及该治疗关系的具体语句或段落引用。方剂节点的“组成药材”属性可以进一步细化,包括每种药材在方剂中的用量、炮制要求等信息。人物节点的“主要贡献”属性可详细列举其在中医药领域的具体学术成果、著作等。

[0105] 步骤204,基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建。

[0106] 在一种可能的实施方式中,所述基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,请参考图7,具体表现为:

[0107] 步骤71,识别出不同中药古籍内部知识图谱的关键实体,所述识别出不同中药古籍内部知识图谱的关键实体,可以通过依据中药古籍的语言特点和专业词汇规律制定识别规则,也可以通过机器学习或者深度学习的方式对不同中药古籍内部知识图谱中的关键实体进行识别。

[0108] 步骤72,获取已识别关键实体之间的语义关系。所述获取已识别关键实体之间的语义关系,可以通过分析中药古籍的语言表达方式,总结出不同实体间关系的常见表述模式,制定相应的抽取规则。比如,若在文本中出现“某某药治某某病”的句式,可确定药材与病症之间存在“治疗”关系;若有“某某方由某某药组成”的表述,则可抽取方剂与药材之间的“组成”关系。通过在知识图谱文本中匹配这些规则模式来抽取实体间的关系。

[0109] 步骤73,基于识别的关键实体以及关键实体之间的语义关系,将每本中药古籍视为关联网络中的一个节点,把不同古籍知识图谱中的关键实体也作为节点,通过对关键实体添加古籍来源属性区分不同古籍中的同名实体,确保网络中节点的唯一性和可识别性;依据语义关系定义边,当两个关键实体来自不同古籍且存在语义关系,则在对应的古籍节点以及关键实体节点之间建立边,以此构建不同中药古籍之间的关联网络。

[0110] 在一种可能的实施方式中,所述基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建,还包括:

[0111] 基于构建的关联网络和已有的知识图谱,从一个关键实体节点出发,通过深度优先搜索与关键节点相关的其他关键实体节点,分析关键节点与其他关键实体节点之间的关系和路径,获取关键节点与其他关键实体节点潜在的关联关系。

[0112] 步骤205,提取现代生物学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建。

[0113] 在一种可能的实施方式中,所述提取现代生物学数据中的疾病病症特征,获取与疾病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建,请参考如图8,具体表现为:

[0114] 步骤81,基于现代生物学数据,采集与疾病病症相关的数据;

[0115] 步骤82,将所述与疾病病症相关的数据作为预设疾病特征模型的输入,获取疾病病症特征;

[0116] 步骤83,基于预设术语映射表,获取疾病病症特征与中药古籍中的相关表述,将所述相关表述作为检索条件,获取与所述疾病病症特征相关的中药古籍内容节点。例如,若疾

病病症特征节点为“感冒”，通过治疗关系、相似病症关系等关联关系，在知识图谱中查找可能与治疗“感冒”相关的中药古籍内容节点。

[0117] 步骤84,提取中药古籍内容节点的相关信息,作为与疾病病症特征关联的中药古籍关联内容。所述相关信息包括对应的古籍书名、章节、具体内容描述等,作为与疾病病症特征关联的中药古籍关联内容。

[0118] 步骤85,利用词向量模型(如Word2Vec、Bert等)将疾病病症特征描述和中药古籍内容分别转化为词向量序列,然后计算两者之间的余弦相似度语义相似度指标,根据设定的相似度阈值(如0.6),判断语义是否匹配。例如,将现代医学中“感冒”的症状描述和治疗方法等转化为词向量,与中药古籍中关于“伤风”的内容词向量进行比较,若余弦相似度大于0.6,则认为在语义上较为匹配。

[0119] 步骤86,当疾病病症特征与中药古籍内容满足语义匹配后,将现代医学中的症状进行编码处理,将古籍中的症状描述通过自然语言处理技术转化为词向量,通过计算疾病病症特征与中药古籍内容之间的相似性指标,判断症状的相似性程度。例如,现代医学中“感冒”的症状“发热、咳嗽、流涕”与古籍中描述的“伤风”症状“发热、咳嗽、鼻流清涕”在经过处理后计算出的相似性指标较高,则认为在症状特征方面较为匹配;

[0120] 步骤87,当疾病病症特征与中药古籍内容症状的相似性程度满足预设阈值时,将疾病病症特征与中药古籍内容在知识图谱层面上实现关联构建,形成一个跨领域的知识网络。

[0121] 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,该计算机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,可包括如上述各方法的实施例的流程。其中,前述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory,ROM)等非易失性存储介质,或随机存储记忆体(Random Access Memory,RAM)等。

[0122] 应该理解的是,虽然附图的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,其可以以其他的顺序执行。而且,附图的流程图中的至少一部分步骤可以包括多个子步骤或者多个阶段,这些子步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,其执行顺序也不必然是依次进行,而是可以与其他步骤或者其他步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

[0123] 继续参考图9,本实施例所述的中药古籍的关联构建装置包括:

[0124] 标准化表达文本获取模块901,用于将中药古籍转化为可编辑的数字文本,基于所述可编辑的数字文本,获取中药古籍的标准化表达文本;

[0125] 关键实体关系确定模块902,用于基于所述中药古籍的标准化表达文本,获取中药古籍的语义关系、上下文信息,确定中药古籍关键实体之间的关系;

[0126] 知识图谱构建模块903,用于基于关键实体之间的关系,构建中药古籍内部的知识图谱;

[0127] 第一关联构建模块904,用于基于所述中药古籍内部的知识图谱,获取不同中药古籍关键实体的关联关系,完成对不同中药古籍的关联构建;

[0128] 关联构建模块模块905,用于提取现代生物医学数据中的疾病病症特征,获取与疾

病病症特征关联的中药古籍关联内容,实现疾病病症特征与中药古籍内容的关联构建。

[0129] 为解决上述技术问题,本申请实施例还提供计算机设备。具体请参阅图6,图6为本实施例计算机设备基本结构框图。

[0130] 所述计算机设备6包括通过系统总线相互通信连接存储器6a、处理器6b、网络接口6c。需要指出的是,图中仅示出了具有组件6a-6c的计算机设备6,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。其中,本技术领域技术人员可以理解,这里的计算机设备是一种能够按照事先设定或存储的指令,自动进行数值计算和/或信息处理的设备,其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit,ASIC)、可编程门阵列(Field-Programmable GateArray,FPGA)、数字处理器(Digital Signal Processor,DSP)、嵌入式设备等。

[0131] 所述计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

[0132] 所述存储器6a至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如,SD或DX存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中,所述存储器6a可以是所述计算机设备6的内部存储单元,例如该计算机设备6的硬盘或内存。在另一些实施例中,所述存储器6a也可以是所述计算机设备6的外部存储设备,例如该计算机设备6上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。当然,所述存储器6a还可以既包括所述计算机设备6的内部存储单元也包括其外部存储设备。本实施例中,所述存储器6a通常用于存储安装于所述计算机设备6的操作系统和各类应用软件,例如中药古籍的关联构建方法的程序代码等。此外,所述存储器6a还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0133] 所述处理器6b在一些实施例中可以是中央处理器(Central Processing Unit,CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器6b通常用于控制所述计算机设备6的总体操作。本实施例中,所述处理器6b用于运行所述存储器6a中存储的程序代码或者处理数据,例如运行所述中药古籍的关联构建方法的程序代码。

[0134] 所述网络接口6c可包括无线网络接口或有线网络接口,该网络接口6c通常用于在所述计算机设备6与其他电子设备之间建立通信连接。

[0135] 本申请还提供了另一种实施方式,即提供一种非易失性计算机可读存储介质,所述非易失性计算机可读存储介质存储有中药古籍的关联构建方法的程序,所述中药古籍的关联构建可被至少一个处理器执行,以使所述至少一个处理器执行如上述的中药古籍的关联构建方法的步骤。

[0136] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服

务器,空调器,或者网络设备等)执行本申请各个实施例所述的方法。

[0137] 显然,以上所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例,附图中给出了本申请的较佳实施例,但并不限制本申请的专利范围。本申请可以以许多不同的形式来实现,相反地,提供这些实施例的目的是使对本申请的公开内容的理解更加透彻全面。尽管参照前述实施例对本申请进行了详细的说明,对于本领域的技术人员来而言,其依然可以对前述各具体实施方式所记载的技术方案进行修改,或者对其中部分技术特征进行等效替换。凡是利用本申请说明书及附图内容所做的等效结构,直接或间接运用在其他相关的技术领域,均同理在本申请专利保护范围之内。

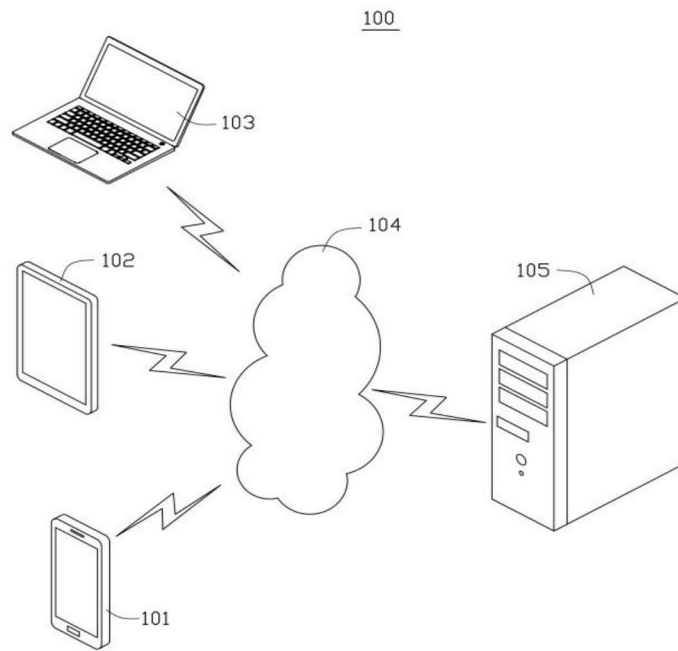


图1

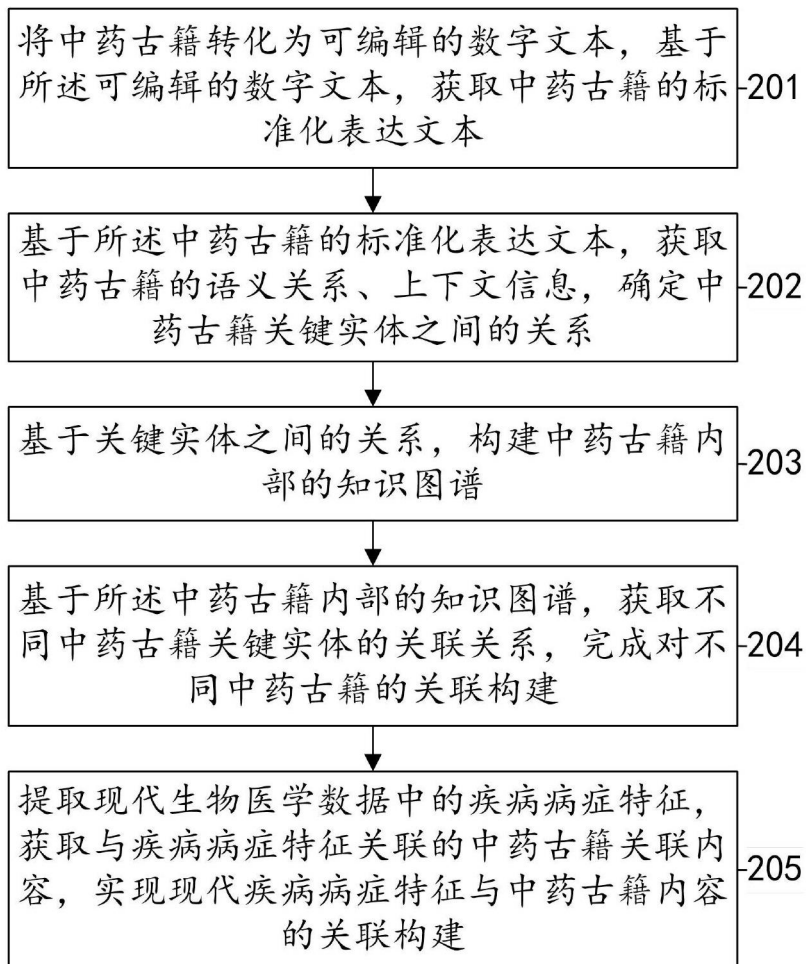


图2

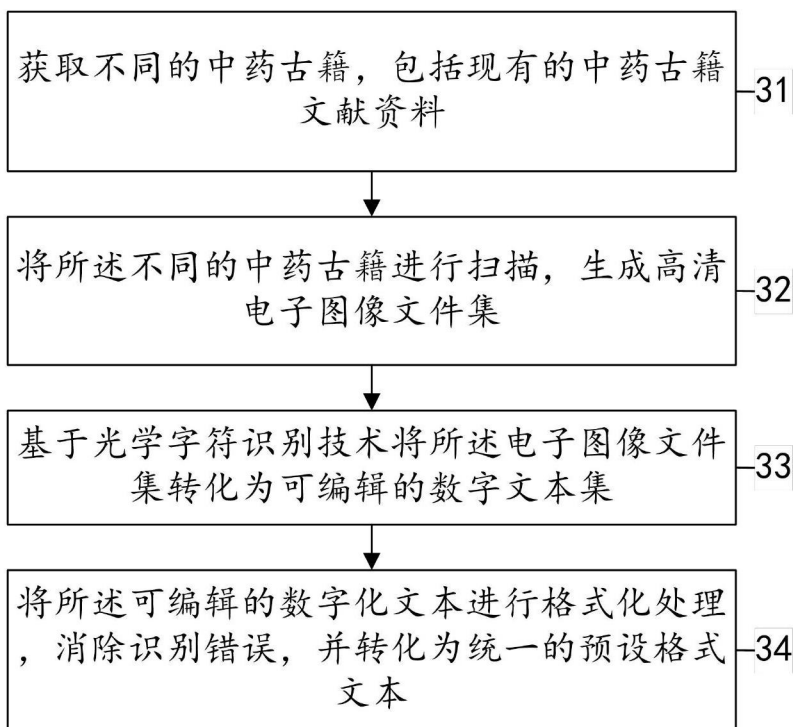


图3

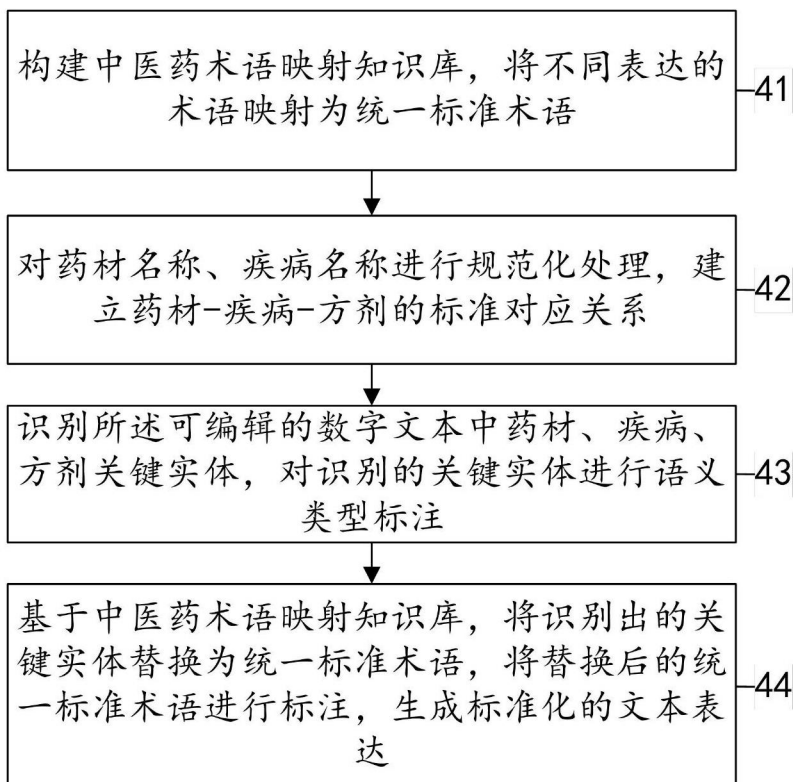


图4

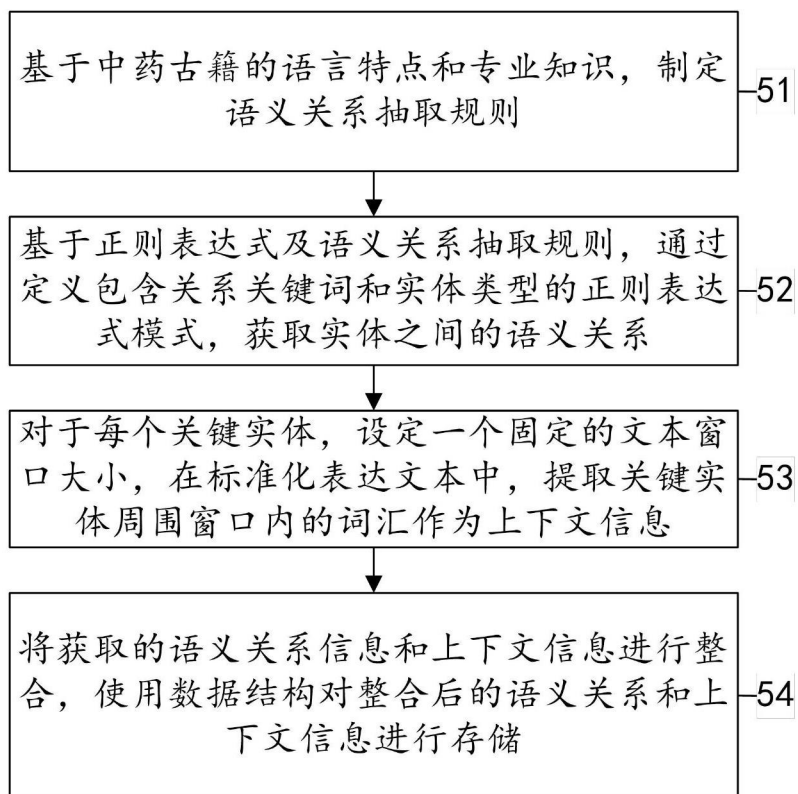


图5

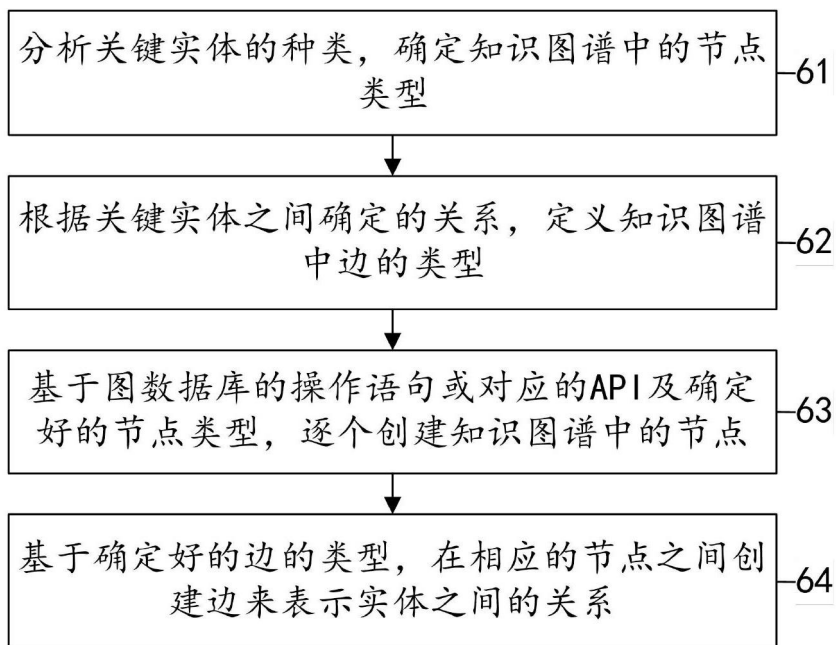


图6

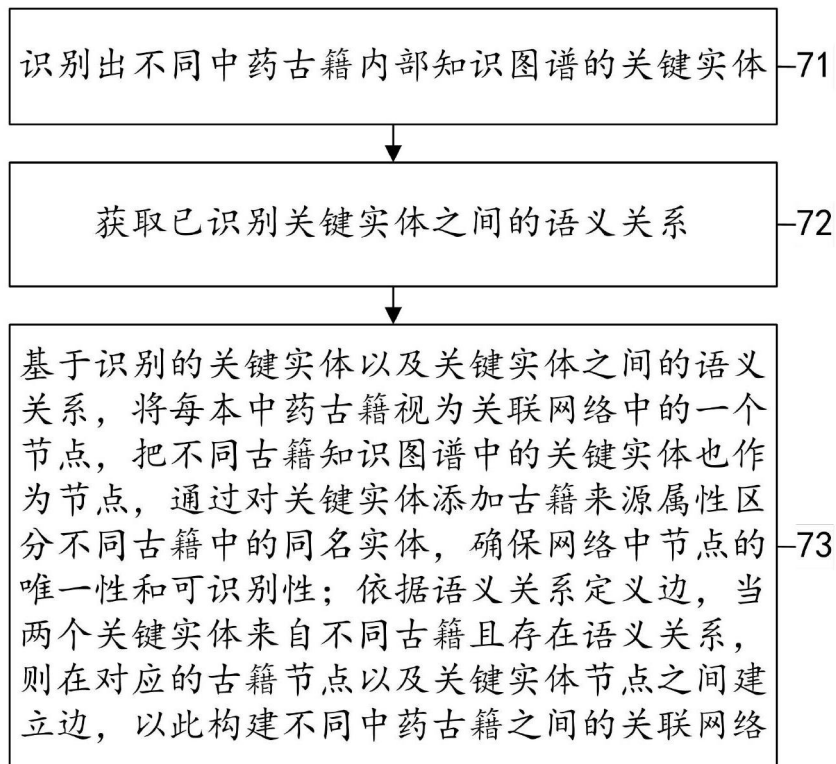


图7

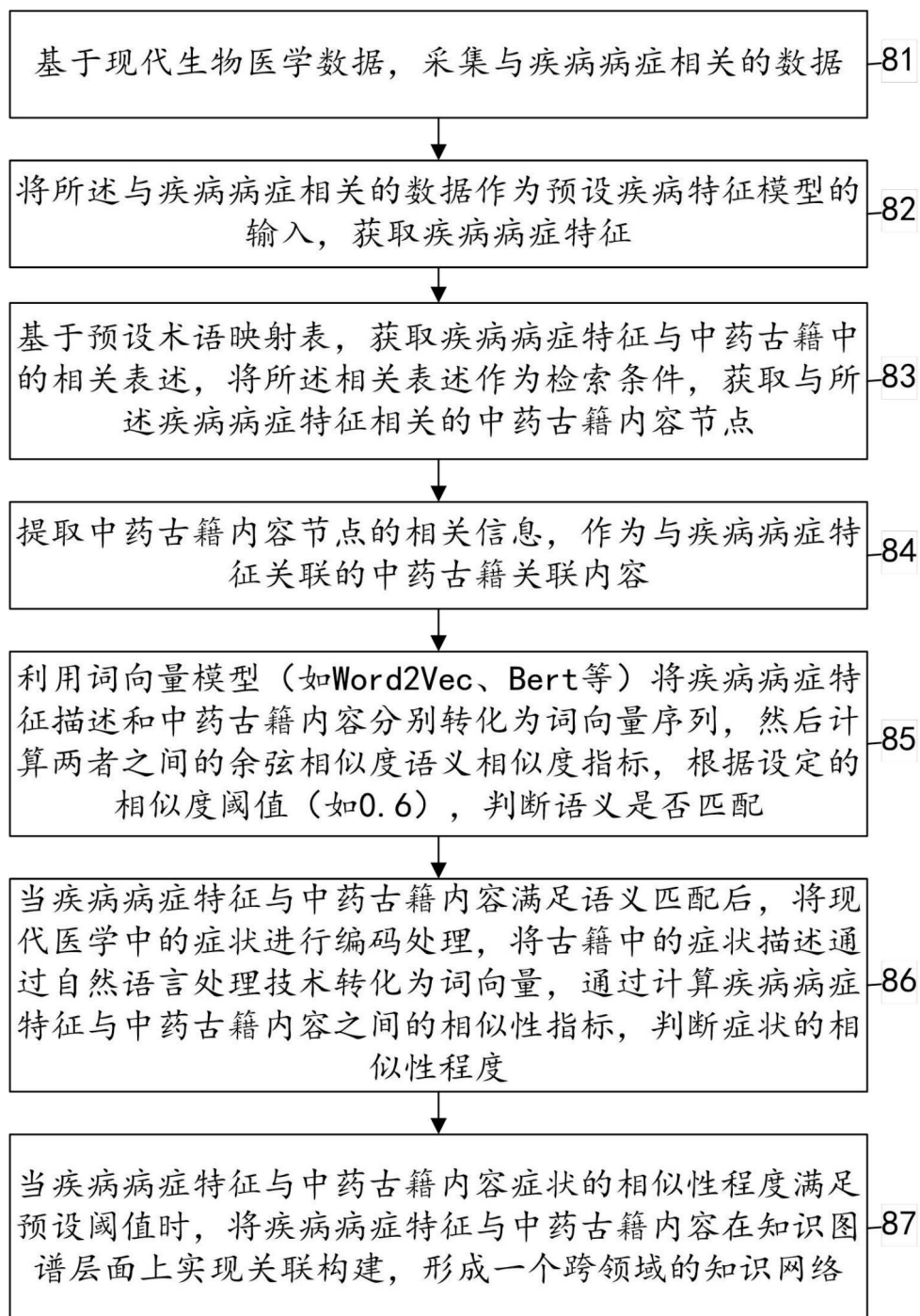


图8

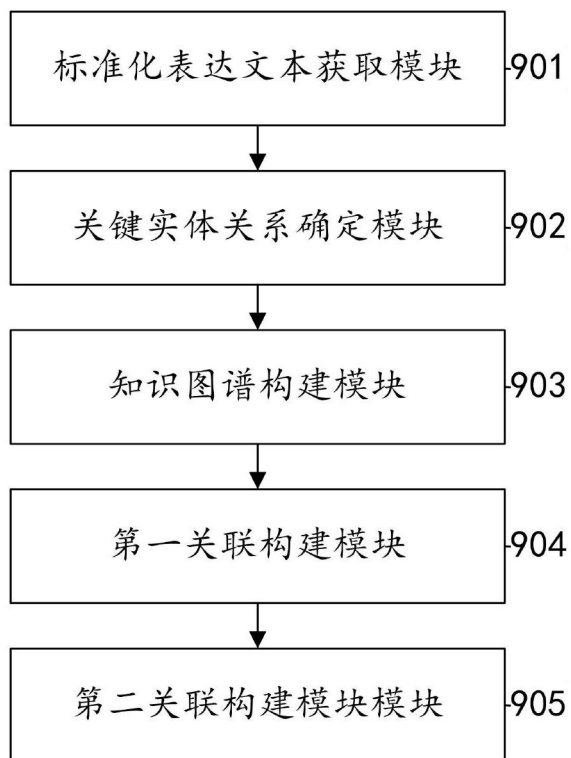


图9

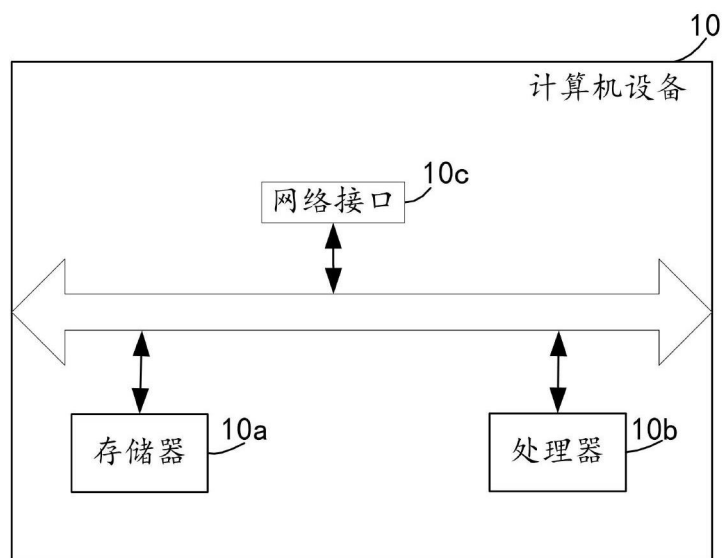


图10