

确定药品编码的方法、装置、电子设备以及计算机介质

相关申请的交叉引用

本申请要求于 2021 年 1 月 15 日提交的申请号为 202110054078.1、发明名称为“确定药品编码的方法、装置、电子设备以及计算机介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本公开涉及计算机技术领域，具体涉及人工智能技术领域，尤其涉及确定药品编码的方法、装置、电子设备、计算机可读介质以及计算机程序产品。

背景技术

解剖学治疗学及化学分类系统，简称 ATC(Anatomical Therapeutic Chemical)系统，是世界卫生组织对药品的官方分类系统。随着医疗信息化系统的建设进展，各级医疗结构、医保局和医疗保险机构，逐步建立起以 ATC 编码体系为基础的药物精确化管理系统。

目前一般通过学习分类算法对药物分子式结构进行预测，得到药物的 ATC 编码。但是通过分子式结构预测药物的 ATC 编码，技术复杂且准确率不高，也不适用于新研发药物之外的药物。

发明内容

本公开的实施例提出了确定药品编码的方法、装置、电子设备、计算机可读介质以及计算机程序产品。

第一方面，本公开的实施例提供了一种确定药品编码的方法，该方法包括：获取药品的说明书文本；提取说明书文本中的药品关键信息；基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分；基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述药品关键信息包括药品成分；，上述基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码，包括：针对至少一个编码中的每个编码，检

测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个，多个规则基于药品成分确定；响应于确定该编码对应的成分满足多个规则中的一个且所有编码均检测完成，确定包括该编码的初筛候选编码；响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述多个规则按从高至低优先级排序如下：1) 药品成分具有两种或以上时，该编码对应的成分包括药品的所有药品成分；2) 药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且含有复方字样；3) 药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且不含有复方字样；4) 药品成分具有一种时，该编码对应的成分包括药品成分。

在一些实施例中，上述药品关键信息包括药品成分，上述基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码，包括：针对至少一个编码中的每个编码，检测该编码对应的成分是否与药品成分相匹配；响应于确定该编码对应的成分与药品成分相匹配且所有编码均检测完成，得到包括该编码的初筛候选编码；响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述药品关键信息还包括药品适应症，上述基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码，还包括：响应于检测到初筛候选编码为多个编码，基于药品适应症，确定药品对应的疾病类型；从初筛候选编码中筛选出与疾病类型对应的编码作为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述基于药品适应症，确定药品对应的疾病类型，包括：采用预先训练完成的分类模型对适应症进行疾病分类，得到分类模型输出的疾病类型。

第二方面，本公开的实施例提供了一种确定药品编码的装置，该装置包括：获取单元，被配置成获取药品的说明书文本；提取单元，被配置成提取说明书文本中的药品关键信息；得到单元，被配置成基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分；筛选单元，被配置成基于药品关键信息以及各个编码对应的成分，对至

少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述药品关键信息包括药品成分，上述筛选单元包括：检测模块，被配置成针对至少一个编码中的每个编码，检测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个，多个规则基于药品成分确定；初筛模块，被配置成响应于确定该编码对应的成分满足多个规则中的一个且所有编码均检测完成，确定包括该编码的初筛候选编码；确定模块，被配置成响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述多个规则按从高至低优先级排序如下：1) 药品成分具有两种或以上时，该编码对应的成分包括药品的所有药品成分；2) 药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且含有复方字样；3) 药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且不含有复方字样；4) 药品成分具有一种时，该编码对应的成分包括药品成分。

在一些实施例中，上述药品关键信息包括药品成分，上述筛选单元包括：匹配模块，被配置成针对至少一个编码中的每个编码，检测该编码对应的成分是否与药品成分相匹配；响应模块，被配置成响应于确定该编码对应的成分与药品成分相匹配且所有编码均检测完成，得到包括该编码的初筛候选编码；编码模块，被配置成响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述药品关键信息还包括药品适应症，筛选单元还包括：分类模块，被配置成响应于检测到初筛候选编码为多个编码，基于药品适应症，确定药品对应的疾病类型；确认模块，被配置成从初筛候选编码中筛选出与疾病类型对应的编码作为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述分类模块进一步地被配置成采用预先训练完成的分类模型对适应症进行疾病分类，得到分类模型输出的疾病类型。

第三方面，本公开的实施例提供了一种电子设备，该电子设备包括：一个或多个处理器；存储装置，其上存储有一个或多个程序；当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

第四方面，本公开的实施例提供了一种计算机可读介质，其上存储有计

计算机程序，该程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

第五方面，本公开的实施例提供了一种计算机程序产品，包括计算机程序，计算机程序在被处理器执行时实现如第一方面任一实现方式描述的方法。

根据本公开的实施例提供的确定药品编码的方法和装置：首先，获取药品的说明书文本；其次，提取说明书文本中的药品关键信息；然后，基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分；最后，基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。由此，能够依据药品的说明书文本，通过预先创建的编码倒排索引，自动化地对药物进行 ATC 编码，解决广大药师在工作中的难题，为医药的信息化系统提供了编码基础信息。

附图说明

通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本公开的其它特征、目的和优点将会变得更明显：

图 1 是本公开的一个实施例可以应用于其中的示例性系统架构图；

图 2 是根据本公开的确定药品编码的方法的一个实施例的流程图；

图 3 是根据本公开的得到药品的解剖学治疗学及化学分类系统编码的方法的一个实施例的流程图；

图 4 是根据本公开的得到药品的解剖学治疗学及化学分类系统编码的方法的另一个实施例的流程图；

图 5 是根据本公开的确定药品编码的装置的实施例的结构示意图；

图 6 是适于用来实现本公开的实施例的电子设备的结构示意图。

具体实施方式

下面结合附图和实施例对本公开作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本公开中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本公开。

图 1 示出了可以应用本公开的确定药品编码的方法的示例性系统架构 100。

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，通常可以包括无线通信链路等等。

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，例如即时通信工具、邮箱客户端等。

终端设备 101、102、103 可以是硬件，也可以是软件；当终端设备 101、102、103 为硬件时，可以是具有通信和控制功能的用户设备，上述用户设备可与服务器 105 进行通信。当终端设备 101、102、103 为软件时，可以安装在上述用户设备中；终端设备 101、102、103 可以实现成多个软件或软件模块（例如用来提供分布式服务的软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

服务器 105 可以是提供各种服务的服务器，例如为终端设备 101、102、103 上药品处理系统提供支持的确定药品编码的后台服务器。后台服务器可以对网络中药品的说明书文本进行分析处理，并将处理结果（如确定的 ATC 编码）反馈给终端设备。

需要说明的是，服务器可以是硬件，也可以是软件。当服务器为硬件时，可以实现成多个服务器组成的分布式服务器集群，也可以实现成单个服务器。当服务器为软件时，可以实现成多个软件或软件模块（例如用来提供分布式服务的软件或软件模块），也可以实现成单个软件或软件模块。在此不做具体限定。

需要说明的是，本公开的实施例所提供的确定药品编码的方法一般由服务器 105 执行。

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。

在本实施例的一些可选实现方式中，如图 2，示出了根据本公开的确定药品编码的方法的一个实施例的流程 200，该确定药品编码的方法包括以下

步骤:

步骤 201, 获取药品的说明书文本。

本实施例中, 药品的说明书是指载明药品的重要信息的法定文件, 是选用药品的法定指南, 用药前准确阅读和理解说明书是安全用药的前提。药品的说明书包括药品的名称、规格、生产企业、有效期、用法、用量、药品成分、适应症或功能主治、禁忌、不良反应和注意事项。其中, 药品的品名包括: 通用名、商品名、英文名、化学名等。使用者一般只要能清楚药品的通用名, 就能避免重复用药。药品的说明书文本是用于表示药品的说明书的内容的文本。

确定药品编码的方法运行于其上的执行主体可以通过各种手段获得说明书文本, 比如, 实时从终端获取说明书文本, 或者从内存中读取说明书文本, 对此本实施例不做限定。

步骤 202, 提取说明书文本中的药品关键信息。

本实施例中, 通过获取药品的说明书文本, 可以对说明书文本进行自然语言处理, 得到药品的关键信息。药品的关键信息包括药品成分或与药品成分相关的信息, 而与药品成分相关的信息包括: 药品的品名、药品的适应症或功能主治、禁忌、不良反应等。本实施例中, 药品成分也可以是药品的主要的成分。

自然语言处理技术目前已经广泛应用于生活中需要语义理解的场景。如实体识别技术, 可以将一段文本中的实体(如药物名称、疾病名称、治疗方法等)识别出来, 这样就可以做到自动分析诸如医生医嘱中诊断、处方等内容, 以结构化方式进行医疗信息管理。如文本分类技术, 可以应用于智能分诊场景, 智能解析患者病情描述, 基于病情描述信息, 精准匹配诊室, 提高分诊效率。自然语言处理技术与医疗场景相结合, 能够提高医疗场景的智能化, 为用户提供更好体验。

本实施例中, 通过自然语言处理可以提取说明书文本中的药品成分、药品的品名、药品的适应症或功能主治、禁忌、不良反应等。在药品说明书中, 药品成分一般包括在一小段文本描述的自然语言中, 例如: 本品为复方制剂, 每毫升含盐酸克林霉素(以克林霉素计)10 毫克, 甲硝唑 8 毫克。辅料为: 甘油、乙醇。借助自然语言处理模型(例如命名实体识别模型), 可以对其中的主要成分(非辅助成分或辅料)进行提取, 针对上述说明书文本, 通过自

然语言处理模型提取的药品成分包括：盐酸克林霉素、甲硝唑、甘油、乙醇。

可选地，可以采用 BERT（Bidirectional Encoder Representation from Transformers，基于多层双向转换解码）+CRF（conditional random field，条件随机场）组成的自然语言模型进行训练，得到训练完成的针对药品成分进行实体识别的命名实体识别模型。通过该命名实体识别模型得到药品关键信息。在实际操作中，该命名实体识别模型的识别结果的准确率可以接近 90%，完全可以满足实际临床的使用要求。

基于药品的性质、特点不同，药品包括复方药和单方药。单方药是指单味药制剂，单方药主要包含一种药品成分。复方药是指两种或两种以上的药物混合制剂，可以是中药、西药或中西药混合。复方药包含两种或以上药品成分。本实施例中，针对上述不同种类的药品，药品关键信息中的药品成分可以指一种，也可以是多种。

步骤 203，基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分。

本实施例中，编码倒排索引是在提取说明书文本中的药品关键信息之前创建的索引库，该编码倒排索引只需创建一次，便可以重复使用。

本实施例中，创建的编码倒排索引基于需要确定的药品的编码确定，本申请提供的确定药品的编码的方法用于确定药品的 ATC 编码，因此编码倒排索引可以是以世界卫生组织所定义的 ATC 编码分类标准分类信息（ATC 中文名称、ATC 英文名称、ATC 编码），按组进行倒排索引，例如一种编码倒排索引如表 1 所示。在表 1 中包括 ATC 编码、ATC 编码对应的化学物质的中文名称、英文名称，其中，化学物质也是药品成分，即 ATC 编码对应的药品成分。例如，“替克拉酮”对应的英文名称是“ticlatone”，对应的 ATC 编码为“D01AE08”。

需要说明的是，由于药品成分可以是一种也可以是多种，针对多种药品成分，其对应的 ATC 编码一定是多个，而针对一种药品成分，其对应的 ATC 编码可以是多个，例如，表 1 中，含“替加氟”的药物的 ATC 编码可以对应的 ATC 编码包括：“L01BC03”、“L01BC53”。

表 1

中文名称	英文名称	ATC 编码
替克拉酮	ticlatone	D01AE08
替克洛可	teclozan	P01AC04
替利定	tilidine	N02AX01
替加氟 (喃氟啉)	tegafur	L01BC03
替加氟, 复方	tegafur, combinations	L01BC53
...	...	...

在实际应用场景中, 可以使用搜索引擎软件 (例如 Elasticsearch) 完成世界卫生组织所定义的 ATC 编码的索引。通过建立倒排索引搜索引擎, 在查找相应的文本字段时, 如查找 ATC 中文名称中含有某字段 (如甲硝唑) 的分类信息, 就可以很容易将所有在中文名称中出现 “甲硝唑” 的 ATC 编码全部查找出来, 如查找到的结果为: 甲硝唑, A01AB17; 兰索拉唑, 阿莫西林和甲硝唑, A02BD03 等。因此通过搜索引擎软件可以很容易得到编码以及与编码对应的成分。

进一步地, 还可以设置搜索引擎软件返回 ATC 编码的数量。将步骤 202 得到的药品关键信息中的所有药品成分放入编码倒排索引中找到与之相关的编码或编码对应的成分, 每种药品成份可以设置最大返回编码或编码对应的成分数目为 n ($n > 1$), 例如 n 设置为 10。

步骤 204, 基于药品关键信息以及各个编码对应的成分, 对至少一个编码进行筛选, 得到药品的解剖学治疗学及化学分类系统编码。

本实施例中, 可选地, 由于至少一个编码可以是一个, 也可以是多个, 所以在得到至少一个编码之后, 可以首先检测该至少一个编码的数量。当至少一个编码为一个时, 则得到的该编码即为 ATC 编码。当至少一个编码为多于一个的编码时, 则需要基于药品关键信息以及各个编码对应的成分, 对至少一个编码进行筛选, 得到药品的 ATC 编码。

由于直接通过倒排索引的检索得到的编码不一定完全满足药品关键信息中药品成分的要求, 所以在本实施例的一些可选实现方式中, 上述基于药品关键信息以及各个编码对应的成分, 对至少一个编码进行筛选, 得到药品的解剖学治疗学及化学分类系统编码, 包括: 针对至少一个编码中的每个编码, 检测该编码对应的成分是否与药品成分相匹配; 响应于确定该编码对应的成

分与药品成分相匹配且所有编码均检测完成，得到包括该编码的初筛候选编码；响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

本可选实现方式中，药品成分可以是采用不同语言表示。检测药品成分与编码对应的成分是否匹配，可以通过两者内容（中文名称或英文单词）的相似度检测；或者可以通过两者的适用的治疗疾病确定，比如药品成分与编码对应的成分均可以治疗两种以上的相同的疾病，确定两者匹配。当然还可以采用其他方式检测药品成分与编码对应的成分是否匹配，对此不做限定。

本实施例中，初筛候选编码包括至少一个编码中与药品的所有药品成分匹配的所有编码，即至少一个该编码，而该编码对应的成分与药品成分相匹配。

本可选实现方式中，通过将药品的关键信息中的药品成分与至少一个编码中每个编码对应的成分进行匹配，在满足匹配条件时，得到包括该编码的初筛候选编码。在至少一个编码的所有编码均检测完成之后，确定初筛候选编码中编码的数量。而在编码只是一个时，得到药品的 ATC 编码。由此，仅通过与倒排索引结果进行药品成分匹配，便可以得到初筛候选编码，实现简单，操作方便。

在本实施例另一些可选实现方式中，药品关键信息还包括药品适应症；基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码，还包括：响应于检测到初筛候选编码为多个编码，基于药品适应症，确定药品对应的疾病类型；从初筛候选编码中筛选出与疾病类型对应的编码作为药品的解剖学治疗学及化学分类系统编码。

根据本公开的实施例提供的确定药品编码的方法：首先，获取药品的说明书文本；其次，提取说明书文本中的药品关键信息；然后，基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分；最后，基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。由此，能够依据药品的说明书文本，通过预先创建的编码倒排索引，自动化地对药物进行 ATC 编码，解决广大药师在工作中的难题，为医药的信息化系统提供了编码基础信息。

在药品关键信息包括药品成分时，在本实施例的一些可选实现方式中，如图 3，示出了根据本公开的得到药品的解剖学治疗学及化学分类系统编码的方法的一个实施例的流程 300，该得到药品的解剖学治疗学及化学分类系统编码的方法包括以下步骤：

步骤 301，针对至少一个编码中的每个编码，检测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个；若该编码对应的成分满足具有优先级顺序的多个规则中的一个，执行步骤 302。

本实施例中，多个规则基于药品成分确定，在编码对应的成分按规则的优先级顺序满足多个规则中的任一个之后，多个规则中其他的规则可不用再考虑。

具体地，多个规则按从高至低优先级排序如下：1) 药品成分具有两种或以上时，该编码对应的成分包括药品的所有药品成分；2) 药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且含有复方字样；3) 药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且不含有复方字样；4) 药品成分具有一种时，该编码对应的成分包括药品成分。

需要说明的是，上述多个规则中的各个规则的内容、优先级顺序以及数目可以基于药品的说明书文本中的药品成分进行适应性调节。例如，针对说明书文本为单方药品的说明书文本，多个规则可以仅包括上述 1) 与 4)。例如，针对说明书文本为复方药品的说明书文本，多个规则可以仅包括上述 1) -3)。

本可选实现方式中，具有优先级顺序的多个规则可以适用于单方药品和复方药品，且以复方药品为优先考虑对象，提高了编码对应的成分排查的可靠性以及全面性。

步骤 302，检测至少一个编码中所有编码是否均检测完成；若是，执行步骤 303；若至少一个编码没有检测完成，返回执行步骤 301。

本实施例中，该编码是至少一个编码中按顺序排布的各个编码，也是当前编码。在步骤 302 中，若当前编码（该编码）满足多个规则中的一个即会被放入初筛候选编码中。若步骤 302 中，当前编码不满足多个规则中的任一个，则放弃该编码，重新返回步骤 301 中，在至少一个编码中当前编码之后

相邻的编码作为当前编码，并重新检测。

步骤 303，确定包括该编码的初筛候选编码，然后执行步骤 304。

本实施例中，初筛候选编码为初次得到的满足药品说明书文本要求的 ATC 编码，初筛候选编码中各个编码对应的成分满足具有优先级顺序的多个规则中的一个，初筛候选编码可以只有一个编码，也可以有多个编码。

本可选实现方式中，初筛候选编码包括上述至少一个编码中所有满足多个规则中的一个的编码，即至少一个该编码，而该编码对应的成分是否满足具有优先级顺序的多个规则中的一个。

步骤 304，检测初筛候选编码是否只有一个编码；若检测结果为只有一个，执行步骤 305。

本实施例中，针对检测结果为只有一个编码时，确定当前的初筛候选编码就是药品的 ATC 编码，无需再进行后续任何检测。

本可选实施例中，针对初筛候选编码是多个编码的情况，可选地，可以将初筛候选编码中的编码进行相似度匹配，将初筛候选编码中相似度最多的多个初筛候选编码中的一个初筛候选编码作为药品的 ATC 编码。

可选地，针对说明书文本为复方药品的说明书文本，可以将所有的初筛候选编码中初筛候选编码对应成分具有复方字样的编码做为药品的 ATC 编码。针对说明书文本为复方药品的说明书文本，可以将所有的初筛候选编码中初筛候选编码对应成分不具有复方字样的编码做为药品的 ATC 编码。

步骤 305，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

本可选实现方式中，在药品关键信息包括药品成分时，基于药品成分对应确定的多个规则，确定药品的解剖学治疗学及化学分类系统编码，提高了 ATC 编码的确定的可靠性。

当药品关键信息包括药品成分和药品适应症时，在本实施例的一些可选实现方式中，如图 4，示出了根据本公开的得到药品的解剖学治疗学及化学分类系统编码的方法的另一个实施例的流程 400。该得到药品的解剖学治疗学及化学分类系统编码的方法包括以下步骤：

步骤 401，针对至少一个编码中的每个编码，检测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个。若该编码对应的成分满足具有

优先级顺序的多个规则中的一个，执行步骤 402。

步骤 402，检测至少一个编码中所有编码是否均检测完成；若是，执行步骤 403；若至少一个编码没有检测完成，返回执行步骤 401。

步骤 403，确定包括该编码的初筛候选编码，然后执行步骤 404。

步骤 404，检测初筛候选编码是否只有一个编码。若检测结果为只有一个编码，执行步骤 405。若检测结果为初筛候选编码为多个编码，执行步骤 406。

步骤 405，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

应当理解，上述步骤 401-步骤 405 中的操作和特征，分别与步骤 301-步骤 305 中的操作和特征相对应，因此，上述在步骤 301-步骤 305 中对于操作和特征的描述，同样适用于步骤 401-步骤 405，在此不再赘述。

步骤 406，基于药品适应症，确定药品对应的疾病类型，然后执行步骤 407。

本实施例中，可选地，可以预先设置适应症与疾病类型对应关系表，在得到药品适应症之后，基于上述预先设置的适应症与疾病类型对应关系表，可以快速得到与药品适应症对应的疾病类型。

在本实施例的一些可选实现方式中，基于药品适应症，确定药品对应的疾病类型包括：采用预先训练完成的分类模型对适应症进行疾病分类，得到分类模型输出的疾病类型。

在实际应用中，可以采用 BERT 模型构建分类模型，以使分类模型对药品的说明书文本中的适应症进行疾病分类，得到模型输出的不同疾病类型的概率值，比如，做 14 个疾病类型的分类。例如说明书中适应症包括“用于寻常痤疮，也可用于脂溢性皮炎及酒渣鼻、毛囊炎”，做 14 个疾病类型的分类，以确定药物属于治疗哪一类疾病的药物。这些分类是消化系统、代谢系统、血液及造血器官、心血管系统、皮肤病、泌尿生殖系统、性激素、抗感染、抗肿瘤及免疫用药、肌骨骼系统、神经系统、抗寄生虫、呼吸系统、感觉系统，共 14 个疾病类型。这 14 个分类也对应于 ATC 一级分类中的 14 个疾病类型。

针对上述说明书中的适应症“用于寻常痤疮，也可用于脂溢性皮炎及酒渣鼻、毛囊炎”，其中，分类模型针对上述 14 个疾病类型中输出各自的置信

度分数。例如,对上述适应症,分类模型输出分类分数分别为:消化系统(2%)、代谢系统(7%)、血液及造血器官(8%)、心血管系统(5%)、皮肤病(80%)、泌尿生殖系统(1%)、性激素(8%)、抗感染(10%)、抗肿瘤及免疫用药(2%)、肌骨骼系统(2%)、神经系统(2%)、抗寄生虫(2%)、呼吸系统(2%)、感觉系统(2%),则得到结果为上述适应症的疾病类型属于皮肤病。

通过训练一个疾病类型与药品适应症相对应的分类模型,即可通过药物说明书的适应症文本的描述得知该药是用来治疗哪方面疾病的,本实施例中,采用分类模型进行分类的分类准确率可达93%以上。

本可选实现方式中,将说明书文本中提取到的适应症输入到预先训练完成的分类模型,可以得到分类模型输出的疾病类型。进一步,将得到的疾病类型与初筛候选编码中各个编码对应的疾病类型进行对比,可以得到初筛候选编码中的药品对应的优选的ATC编码。通过分类模型可以提高疾病类型获取的准确度,保证了药品的ATC编码得到的可靠性。

步骤407,从初筛候选编码中筛选出与疾病类型对应的编码作为药品的解剖学治疗学及化学分类系统编码。

本实施例中,疾病类型可以是一个,也可以是多个;在疾病类型为一个时,与疾病类型对应的初筛候选编码即为药品的ATC编码。在疾病类型为多个时,可选地,可以将与该疾病类型中最多疾病类型对应的初筛候选编码做为药品的ATC编码。当然还可以选取得到的第一个疾病类型对应的编码最为药品的ATC编码。对此,本申请不做限定。

本可选实现方式中,在药品关键信息包括药品成分和药品适应症时,基于药品成分确定至少一个编码中的初筛候选编码,在初筛候选编码为多个编码时,基于药品适应症确定药品对应的疾病类型,从初筛候选编码中确定药品的ATC编码,解决了同一个药品具有多个ATC编码的问题,保证了ATC编码确定的准确性。

进一步参考图5,作为对上述各图所示方法的实现,本公开提供了确定药品编码的装置的一个实施例。该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

如图5所示,本公开的实施例提供了一种确定药品编码的装置500,该装置500包括:获取单元501、提取单元502、得到单元503、筛选单元504。

其中，获取单元 501，可以被配置成被配置成获取药品的说明书文本。提取单元 502，可以被配置成提取说明书文本中的药品关键信息。得到单元 503，可以被配置成基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分。筛选单元 504，可以被配置成基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。

在本实施例中，确定药品编码的装置 500 中，获取单元 501、提取单元 502、得到单元 503、筛选单元 504 的具体处理及其所带来的技术效果可分别参考图 2 对应实施例中的步骤 201、步骤 202、步骤 203、步骤 204。

在一些实施例中，上述药品关键信息包括药品成分，上述筛选单元 504 包括：检测模块（图中未示出）、初筛模块（图中未示出）、确定模块（图中未示出）。其中，检测模块，可以被配置成针对至少一个编码中的每个编码，检测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个，多个规则基于药品成分确定。初筛模块，可以被配置成响应于确定该编码对应的成分满足多个规则中的一个且所有编码均检测完成，确定包括该编码的初筛候选编码。确定模块，可以被配置成响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述多个规则按从高至低优先级排序如下：1）药品成分具有两种或以上时，该编码对应的成分包括药品的所有药品成分；2）药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且含有复方字样；3）药品成分具有两种或以上时，该编码对应的成分包括药品的至少一个药品成分且不含有复方字样；4）药品成分具有一种时，该编码对应的成分包括药品成分。

在一些实施例中，上述药品关键信息包括药品成分，上述筛选单元 504 包括：匹配模块（图中未示出）、响应模块（图中未示出）、编码模块（图中未示出）。其中，匹配模块，可以被配置成针对至少一个编码中的每个编码，检测该编码对应的成分是否与药品成分相匹配。响应模块，可以被配置成响应于确定该编码对应的成分与药品成分相匹配且所有编码均检测完成，得到包括该编码的初筛候选编码。编码模块，可以被配置成响应于检测到初筛候选编码只有一个编码，确定初筛候选编码为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述药品关键信息还包括：药品适应症；上述筛选单元 504 包括：分类模块（图中未示出）、确认模块（图中未示出）。其中，分类模块，可以被配置成响应于检测到初筛候选编码为多个编码，基于药品适应症，确定药品对应的疾病类型。上述确认模块，可以被配置成从初筛候选编码中筛选出与疾病类型对应的编码作为药品的解剖学治疗学及化学分类系统编码。

在一些实施例中，上述分类模块进一步地被配置成采用预先训练完成的分类模型对适应症进行疾病分类，得到分类模型输出的疾病类型。

根据本公开的实施例提供的确定药品编码的方法：首先，获取单元 501 获取药品的说明书文本；其次，提取单元 502 提取说明书文本中的药品关键信息；然后，得到单元 503 基于预先创建的编码倒排索引，得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分；最后，筛选单元 504 基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。由此，能够依据药品的说明书文本，通过预先创建的编码倒排索引，自动化地对药物进行 ATC 编码，解决广大药师在工作中的难题，为医药的信息化系统提供了编码基础信息。

下面参考图 6，其示出了适于用来实现本公开的实施例的电子设备 600 的结构示意图。

如图 6 所示，电子设备 600 可以包括处理装置（例如中央处理器、图形处理器等）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储装置 608 加载到随机访问存储器（RAM）603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有电子设备 600 操作所需的各种程序和数据。处理装置 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

通常，以下装置可以连接至 I/O 接口 605：包括例如触摸屏、触摸板、键盘、鼠标等的输入装置 606；包括例如液晶显示器（LCD，Liquid Crystal Display）、扬声器、振动器等的输出装置 607；包括例如磁带、硬盘等的存储装置 608；以及通信装置 609。通信装置 609 可以允许电子设备 600 与其他设备进行无线或有线通信以交换数据。虽然图 6 示出了具有各种装置的电子设备 600，但是应理解的是，并不要求实施或具备所有示出的装置。可以替

代地实施或具备更多或更少的装置。图 6 中示出的每个方框可以代表一个装置，也可以根据需要代表多个装置。

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程序产品，其包括承载在计算机可读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装置 609 从网络上被下载和安装，或者从存储装置 608 被安装，或者从 ROM 602 被安装。在该计算机程序被处理装置 601 执行时，执行本公开的实施例的方法中限定的上述功能。

需要说明的是，本公开的实施例的计算机可读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。在本公开的实施例中，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。而在本公开的实施例中，计算机可读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介质，该计算机可读信号介质可以发送、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：电线、光缆、RF（Radio Frequency，射频）等等，或者上述的任意合适的组合。

上述计算机可读介质可以是上述服务器中所包含的；也可以是单独存在，而未装配入该服务器中。上述计算机可读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该服务器执行时，使得该服务器：获取药品的说明书文本；提取说明书文本中的药品关键信息；基于预先创建的编码倒排索引，

得到与药品关键信息相关的至少一个编码以及与各个编码对应的成分；基于药品关键信息以及各个编码对应的成分，对至少一个编码进行筛选，得到药品的解剖学治疗学及化学分类系统编码。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的实施例的操作的计算机程序代码，程序设计语言包括面向对象的设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开的各种实施例的系统、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，该模块、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。也要注意的，框图和/或流程图中的每个方框、以及框图和/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。

描述于本公开的实施例中所涉及到的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，包括获取单元、提取单元、得到单元以及筛选单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，获取单元还可以被描述为“被配置成获取药品的说明书文本”的单元。

以上描述仅为本公开的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开的实施例中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方

案。例如上述特征与本公开的实施例中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

权利要求书

1、一种确定药品编码的方法，所述方法包括：

获取药品的说明书文本；

提取所述说明书文本中的药品关键信息；

基于预先创建的编码倒排索引，得到与所述药品关键信息相关的至少一个编码以及与各个编码对应的成分；

基于所述药品关键信息以及各个编码对应的成分，对所述至少一个编码进行筛选，得到所述药品的解剖学治疗学及化学分类系统编码。

2、根据权利要求1所述的方法，其中，所述药品关键信息包括：药品成分，

所述基于所述药品关键信息以及各个编码对应的成分，对所述至少一个编码进行筛选，得到所述药品的解剖学治疗学及化学分类系统编码，包括：

针对所述至少一个编码中的每个编码，检测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个，所述多个规则基于所述药品成分确定；

响应于确定该编码对应的成分满足所述多个规则中的一个且所有编码均检测完成，确定包括该编码的初筛候选编码；

响应于检测到所述初筛候选编码只有一个编码，确定所述初筛候选编码为所述药品的解剖学治疗学及化学分类系统编码。

3、根据权利要求2所述的方法，其中，所述多个规则按从高至低优先级排序如下：

1)所述药品成分具有两种或以上时，该编码对应的成分包括所述药品的所有药品成分；

2)所述药品成分具有两种或以上时，该编码对应的成分包括所述药品的至少一个药品成分且含有复方字样；

3)所述药品成分具有两种或以上时，该编码对应的成分包括所述药品的至少一个药品成分且不含有复方字样；

4)所述药品成分具有一种时，该编码对应的成分包括所述药品成分。

4、根据权利要求1所述的方法，其中，所述药品关键信息包括药品成

分，

所述基于所述药品关键信息以及各个编码对应的成分，对所述至少一个编码进行筛选，得到所述药品的解剖学治疗学及化学分类系统编码，包括：

针对所述至少一个编码中的每个编码，检测该编码对应的成分是否与所述药品成分相匹配；

响应于确定该编码对应的成分与所述药品成分相匹配且所有编码均检测完成，得到包括该编码的初筛候选编码；

响应于检测到所述初筛候选编码只有一个编码，确定所述初筛候选编码为所述药品的解剖学治疗学及化学分类系统编码。

5、根据权利要求 2-4 之一所述的方法，其中，所述药品关键信息还包括药品适应症，

所述基于所述药品关键信息以及各个编码对应的成分，对所述至少一个编码进行筛选，得到所述药品的解剖学治疗学及化学分类系统编码，还包括：

响应于检测到所述初筛候选编码为多个编码，基于所述药品适应症，确定所述药品对应的疾病类型；

从所述初筛候选编码中筛选出与所述疾病类型对应的编码作为所述药品的解剖学治疗学及化学分类系统编码。

6、根据权利要求 5 所述的方法，其中，所述基于所述药品适应症，确定所述药品对应的疾病类型，包括：

采用预先训练完成的分类模型对所述适应症进行疾病分类，得到所述分类模型输出的疾病类型。

7、一种确定药品编码的装置，所述装置包括：

获取单元，被配置成获取药品的说明书文本；

提取单元，被配置成提取所述说明书文本中的药品关键信息；

得到单元，被配置成基于预先创建的编码倒排索引，得到与所述药品关键信息相关的至少一个编码以及与各个编码对应的成分；

筛选单元，被配置成基于所述药品关键信息以及各个编码对应的成分，对所述至少一个编码进行筛选，得到所述药品的解剖学治疗学及化学分类系

统编码。

8、根据权利要求 7 所述的装置，其中，所述药品关键信息包括药品成分，所述筛选单元包括：

检测模块，被配置成针对所述至少一个编码中的每个编码，检测该编码对应的成分是否满足具有优先级顺序的多个规则中的一个，所述多个规则基于所述药品成分确定；

初筛模块，被配置成响应于确定该编码对应的成分满足所述多个规则中的一个且所有编码均检测完成，确定包括该编码的初筛候选编码；

确定模块，被配置成响应于检测到所述初筛候选编码只有一个编码，确定所述初筛候选编码为所述药品的解剖学治疗学及化学分类系统编码。

9、根据权利要求 8 所述的装置，其中，所述药品关键信息还包括药品适应症，

所述筛选单元还包括：

分类模块，被配置成响应于检测到所述初筛候选编码为多个编码，基于所述药品适应症，确定所述药品对应的疾病类型；

确认模块，被配置成从所述初筛候选编码中筛选出与所述疾病类型对应的编码作为所述药品的解剖学治疗学及化学分类系统编码。

10、一种电子设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器实现如权利要求 1-6 中任一所述的方法。

11、一种计算机可读介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求 1-6 中任一所述的方法。

12、一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现如权利要求 1-6 中任一项所述的方法。