

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2021 年 9 月 23 日 (23.09.2021)



(10) 国际公布号
WO 2021/184552 A1

(51) 国际专利分类号:
G06F 16/33 (2019.01)

(21) 国际申请号: PCT/CN2020/093322

(22) 国际申请日: 2020 年 5 月 29 日 (29.05.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010195666.2 2020年3月19日 (19.03.2020) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 吴汇哲(WU, Huizhe); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳众鼎专利商标代理事务所(普通合伙)(SHENZHEN ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区笋岗东路 3012 号中民时代广场 B 座 7 楼 701 室, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) Title: MEDICAL TEXT SEARCH METHOD AND APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 医疗文本搜索方法、装置、计算机设备及存储介质

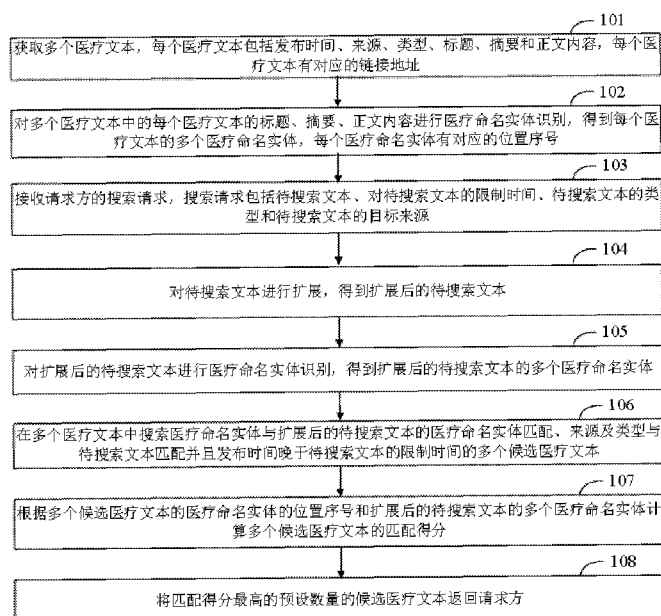


图 1

- 101 Acquire a plurality of medical texts, each medical text comprising publication time, source, type, title, abstract and text content, each medical text having a corresponding link address
- 102 Perform medical named entity recognition on the title, abstract and text content of each of the plurality of medical texts, to obtain a plurality of medical named entities of each medical text, each medical named entity having a corresponding location serial number
- 103 Receive a search request of a requester, the search request comprising a text to be searched for, the limitation time of said text, the type of said text, a target source of said text
- 104 Expand said text, to obtain an expanded text
- 105 Perform medical named entity recognition on the expanded text, to obtain a plurality of medical named entities of the expanded text
- 106 Search in the plurality of medical texts for a plurality of candidate medical texts of which the medical named entities match the medical named entities of the expanded text, of which the source and type match the text to be searched for and of which the publication time is later than the limitation time of the text to be searched for
- 107 Calculate matching scores of the plurality of candidate medical texts according to the location serial numbers of the medical named entities of the plurality of candidate medical texts and the plurality of medical named entities of the expanded text
- 108 Return to the requester a preset number of candidate medical texts having the highest matching score

(57) Abstract: The present application relates to the field of smart medical technology, and provides a medical text search method and a related device. Said method comprises: acquiring a plurality of medical texts; performing medical named entity recognition on each of the plurality of medical texts, to obtain a plurality of medical named entities of each medical text; receiving a search request of a requester; expanding a text to be searched for; performing medical named entity recognition on the expanded text; searching in the plurality of medical texts for a plurality of candidate medical texts of which the medical named entities match the expanded text, of

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

which the source and type match the text to be searched for and of which the publication time is later than the limitation time of the text to be searched for; calculating matching scores of the plurality of candidate medical texts according to the location serial numbers of the medical named entities of the plurality of candidate medical texts and the plurality of medical named entities of the expanded text; and returning to the requester a preset number of candidate medical texts having the highest matching score. The present application improves the accuracy of medical text search.

(57) 摘要: 本申请涉及智慧医疗技术领域, 提供一种医疗文本搜索方法及相关设备。所述方法获取多个医疗文本; 对多个医疗文本中的每个医疗文本进行医疗命名实体识别, 得到每个医疗文本的多个医疗命名实体; 接收请求方的搜索请求; 对待搜索文本进行扩展; 对扩展后的待搜索文本进行医疗命名实体识别; 在多个医疗文本中搜索医疗命名实体与扩展后的待搜索文本匹配、来源及类型与待搜索文本匹配并且发布时间晚于待搜索文本的限制时间的多个候选医疗文本; 根据多个候选医疗文本的医疗命名实体的位置序号和扩展后的待搜索文本的多个医疗命名实体计算多个候选医疗文本的匹配得分; 将匹配得分最高的预设数量的候选医疗文本返回请求方。本申请提升医疗文本搜索的准确度。

医疗文本搜索方法、装置、计算机设备及存储介质

本申请要求于 2020 年 3 月 19 日提交中国专利局、申请号为 202010195666.2，发明名称“医疗文本搜索方法、装置、计算机设备及存储介质”的中国发明专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及智慧医疗技术领域，具体涉及一种医疗文本搜索方法、装置、计算机设备及可读存储介质。

背景技术

蕴含在医疗文本中的知识对生物医学的医疗实践、教学和科研都有重要的意义。而医疗文本的巨大数量使得人们在海量的文献集中发现和获取这些有用的信息变得愈加困难。

技术问题

因此，发明人意识到，针对海量医疗文本的准确的搜索工具成为相关人员的迫切需要。

技术解决方案

鉴于以上内容，有必要提出一种医疗文本搜索方法、装置、计算机设备及可读存储介质，其可以识别待识别药物语句中的药物。

本申请的第一方面提供一种医疗文本搜索方法，所述方法包括：

获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；

接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

本申请的第二方面提供一种医疗文本搜索装置，所述装置包括：

获取模块，用于获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

第一识别模块，用于对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；

接收模块，用于接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

扩展模块，用于对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

第二识别模块，用于对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

搜索模块，用于在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

计算模块，用于根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

返回模块，用于将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

本申请的第三方面提供一种计算机设备，所述计算机设备包括处理器，所述处理器用于执行存储器中存储的计算机可读指令时实现以下步骤：

获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；

接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

本申请的第四方面提供一个或多个存储有计算机可读指令的可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如下步骤：

获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；

接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

有益效果

本申请实现了根据用户的输入从医疗文本集中搜索匹配医疗文本，提升了医疗文本

搜索的准确度。

本申请的一个或多个实施例的细节在下面的附图和描述中提出，本申请的其他特征和优点将从说明书、附图以及权利要求变得明显。

附图说明

图 1 是本申请实施例提供的医疗文本搜索方法的流程图。

图 2 是本申请实施例提供的医疗文本搜索装置的结构图。

图 3 是本申请实施例提供的计算机设备的示意图。

具体实施方式

为了能够更清楚地理解本申请的上述目的、特征和优点，下面结合附图和具体实施例对本申请进行详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本申请，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。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。

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。本文中在本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。

优选地，本申请的医疗文本搜索方法应用在一个或者多个计算机设备中。所述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 Gate Array, FPGA)、数字处理器(Digital Signal Processor, DSP)、嵌入式设备等。

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。

实施例一

图 1 是本申请实施例一提供的医疗文本搜索方法的流程图。所述医疗文本搜索方法应用于计算机设备，用于根据用户的输入从医疗文本集中搜索匹配医疗文本。

如图 1 所示，所述医疗文本搜索方法包括：

101，获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址。

在一具体实施例中，医疗文本可以是网页，可以直接从网络中搜索获取或通过网络爬虫从网络中抓取。每个网页可以包括该网页的发布时间、该网页的来源、该网页的内容的类型、该网页的标题、摘要和正文内容等。

通过网络爬虫技术可以将互联网中数以百亿计的网页信息保存到本地。具体为通过爬虫代码程序模拟浏览器向网络服务器发送医疗文本请求，以便将医疗文本从网络流中读取出来并保存到本地，此外，还进一步基于相关信息提取规则，从爬取的信息中提取需要的信息。优选通过 Docker 容器作为媒介部署指定的爬虫程序，以爬取医疗文本。爬取的网站包括指定的公示网站以及通过搜索引擎搜索到的网站。

该医疗文本的来源可以包括中华医学会、中国医师协会、中国抗癌协会等。该医疗文本的类型可以包括指南、共识、规范等。

该医疗文本对应的链接地址可以表示该医疗文本的存储路径；可以通过该医疗文本对应的链接地址定位并获取该医疗文本。

在另一实施例中，该医疗文本的类型还可以包括内科、外科、妇科等科室类型。

102，对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实

体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号。

在一具体实施例中，所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别包括：

(1) 对每个医疗文本的标题、摘要、正文内容进行分句、去重、错误修正预处理，得到预处理后的文本内容。

可以对每个医疗文本的标题、摘要、正文内容进行分句预处理、去重预处理、错误修正预处理等。

(2) 将所述预处理后的文本内容中的每个语句转化为向量序列。

可以对每个语句进行分词或分字。将每个语句的多个词通过训练好的词嵌入模型将每个词转化为词向量；或将每个语句的多个字通过训练好的字嵌入模型将每个字转化为字向量。依词序组合每个语句的词向量得到该语句的向量序列，或依字序组合每个语句的字向量得到该语句的向量序列。

(3) 将每个语句对应的向量序列输入训练好的第一双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第一条件随机场，得到该语句的标记序列。

所述第一双向长短时记忆网络可以提取该语句的语义信息和该语句中每个词（或字）的上下文信息，并以特征向量序列的形式将提取的信息传递给所述第一条件随机场，通过所述第一条件随机场对该语句中的每个词（或字）进行分类，得到该语句的标记序列。

(4) 根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

医疗命名实体可以包括疾病、症状和药品等。

例如，医疗文本中的一个语句为“糖尿病是危害人类身心健康的主要慢性疾病之一”的标记序列为“B-1 I-1 E-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0”，则该语句的医疗命名实体为疾病“糖尿病”。其中，“B-1”表示疾病（用1表示疾病，2表示症状，3表示药品）的开始字，“I-1”表示疾病的中间字，“E-1”表示疾病的末尾字，“0”表示其他字（非命名实体的字）。

在另一实施例中，在所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别之前，所述方法还包括：

获取医疗训练语句和所述医疗训练语句的标签向量序列；

将所述医疗训练语句转化为向量序列；

将所述医疗训练语句的向量序列输入所述第一双向长短时记忆网络，得到所述医疗训练语句的特征向量序列，将所述医疗训练语句的特征向量序列输入所述第一条件随机场，得到所述医疗训练语句的标记序列；

根据所述医疗训练语句的标签向量序列和标记序列通过梯度下降法调整所述第一双向长短时记忆网络和所述第一条件随机场的隐藏层的权重和偏置值；

循环执行上述步骤，得到所述训练好的第一双向长短时记忆网络和所述训练好的第一条件随机场。

103，接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源。

在一具体实施例中，可以通过对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源对所述待搜索文本进行筛选。例如，待搜索文本为“糖尿病治疗”，对待搜索文本的限制时间为“2019”，待搜索文本的类型为“指南”，待搜索文本的目标来源为“中华医学会”。

104，对所述待搜索文本进行扩展，得到扩展后的待搜索文本。

在一具体实施例中，所述对所述待搜索文本进行扩展包括：

对所述待搜索文本进行分词处理，得到多个目标词语；

确定所述多个词语中每个目标词语的近义词集合；

在所述待搜索文本中用每个目标词语的近义词集合中的词语替换该目标词语，将每次替换后的所述待搜索文本确定为一个扩展后的待搜索文本。

在另一实施例中，所述确定所述多个词语中每个目标词语的近义词集合包括：

从预设近义词词典中查询该目标词语的近义词；和/或

获取历史搜索信息，从所述历史搜索信息中标记的医疗文本中确定该目标词语的近义词；和/或

从所述待搜索文本中删除该目标词语，得到该目标词语的匹配模板，从所述多个医疗文本中查找所述匹配模板，将查找到的所述匹配模板所在的语句中与该目标词语对应的词语确定为该目标词语的近义词，其中，该目标词语在所述匹配模板中的位置与该目标词语对应的词语在查找到的所述匹配模板中的位置一致。

105，对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体。

在一具体实施例中，所述对所述扩展后的待搜索文本进行医疗命名实体识别包括：

对所述扩展后的待搜索文本进行预处理，得到预处理后的扩展后的待搜索文本；

将所述预处理后的扩展后的待搜索文本中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第二双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第二条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

106，在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本。

例如，可以在所述多个医疗文本中搜索医疗命名实体与扩展后的待搜索文本的医疗命名实体（如糖尿病、高血糖等）匹配、来源（如中华医学会）及类型（如指南）与待搜索文本匹配并且发布时间晚于待搜索文本的限制时间（如 2019）的多个候选医疗文本。

107，根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分。

在一具体实施例中，所述根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分包括：

对于所述多个候选医疗文本中的第 i 个医疗文本， $1 \leq i \leq n$ ， n 为所述多个候选医疗文本的数量，根据第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的位置序号确定第 i 个医疗文本的位置权重 W_i ：

$$W_i = \sum_{j=1}^{m_i} D_{i,j}, \quad D_{i,j} = \begin{cases} d_1, O_{i,j} \in E_{i,1} \\ d_2, O_{i,j} \in E_{i,2} \\ d_3, O_{i,j} \in E_{i,3} \\ d_4, O_{i,j} \in E_{i,4} \end{cases}$$

其中， m_i 为第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的数量， $O_{i,j}$ 表示第 i 个医疗文本中第 j 个医疗命名实体的位置序号， $E_{i,1}$ 表示位于第 i 个医疗文本的标题的位置序号的集合， $E_{i,2}$ 表示位于第 i 个医疗文本的摘要的位置序号的集合， $E_{i,3}$ 表示位于第 i 个医疗文本的正文内容的头尾语句的位置序号的集合， $E_{i,4}$ 表示位于第 i 个医疗文本的正文内容的非头尾语句的位置序号的集合， d_1 、 d_2 、 d_3 和 d_4 为预设值；

根据第 i 个医疗文本的医疗命名实体和所述扩展后的待搜索文本的医疗命名实体计算第 i 个医疗文本的长度匹配度 l_i ：

$$l_i = \frac{C_i}{\max(A_i, B)}$$

其中， C_i 表示第 i 个医疗文本中与所述扩展后的待搜索文本的医疗命名实体一致的医疗命名实体的数量， A_i 表示第 i 个医疗文本中医疗命名实体的数量， B 表示所述扩展后的待搜

索文本中医疗命名实体的数量；

从第 i 个医疗文本中查找词语与所述扩展后的待搜索文本的词语一致的目标文本，计算所述目标文本与所述扩展后的待搜索文本的逆序数 G_i ，根据逆序数 G_i 和所述扩展后的待搜索文本计算第 i 个医疗文本的词序相似度 s_i ：

$$s_i = 1 - \frac{G_i}{F - 1}$$

其中， F 表示所述扩展后的待搜索文本中词语的数量；

根据第 i 个医疗文本的位置权重、长度匹配度和词序相似度计算第 i 个医疗文本的匹配得分 P_i ：

$$P_i = W_i(w_{i,1}l_i + w_{i,2}s_i)$$

其中， $w_{i,1}$ 和 $w_{i,2}$ 为预设权重。

具体地， $d_1 > d_2 > d_3 > d_4$ ， d_1 、 d_2 、 d_3 、 d_4 可以分别取值为 4、3、2、1； $w_{i,1}$ 、 $w_{i,2}$ 可以分别取值为 0.5、0.5。

108，将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

在一具体实施例中，可以将匹配得分最高的预设数量的候选医疗文本的摘要和对应的链接地址返回所述请求方。例如，可以将匹配得分最高的 5 个候选医疗文本的摘要和对应的链接地址返回给请求方，使请求方可以通过点击该链接地址获取更详细匹配度更高的信息。

实施例一的医疗文本搜索方法获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；对所述待搜索文本进行扩展，得到扩展后的待搜索文本；对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；将匹配得分最高的预设数量的候选医疗文本返回所述请求方。实施例一根据用户的输入从医疗文本集中搜索匹配医疗文本，提升了医疗文本搜索的准确度。

实施例二

图 2 是本申请实施例二提供的医疗文本搜索装置的结构图。所述医疗文本搜索装置 20 应用于计算机设备。所述医疗文本搜索装置 20 用于根据用户的输入从医疗文本集中搜索匹配医疗文本。

如图 2 所示，所述医疗文本搜索装置 20 可以包括获取模块 201、第一识别模块 202、接收模块 203、扩展模块 204、第二识别模块 205、搜索模块 206、计算模块 207、返回模块 208。

获取模块 201，用于获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址。

在一具体实施例中，医疗文本可以是网页，可以直接从网络中搜索获取或通过网络爬虫从网络中抓取。每个网页可以包括该网页的发布时间、该网页的来源、该网页的内容的类型、该网页的标题、摘要和正文内容等。

通过网络爬虫技术可以将互联网中数以百亿计的网页信息保存到本地。具体为通过爬虫代码程序模拟浏览器向网络服务器发送医疗文本请求，以便将医疗文本从网络流中读取出来并保存到本地，此外，还进一步基于相关信息提取规则，从爬取的信息中提取需要的信息。优选通过 Docker 容器作为媒介部署指定的爬虫程序，以爬取医疗文本。爬取的网站包括指定的公示网站以及通过搜索引擎搜索到的网站。

该医疗文本的来源可以包括中华医学会、中国医师协会、中国抗癌协会等。该医疗文本

的类型可以包括指南、共识、规范等。该医疗文本对应的链接地址可以表示该医疗文本的存储路径；可以通过该医疗文本对应的链接地址定位并获取该医疗文本。

在另一实施例中，该医疗文本的类型还可以包括内科、外科、妇科等科室类型。

第一识别模块 202，用于对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号。

在一具体实施例中，所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别包括：

(1) 对每个医疗文本的标题、摘要、正文内容进行分句、去重、错误修正预处理，得到预处理后的文本内容。

可以对每个医疗文本的标题、摘要、正文内容进行分句预处理、去重预处理、错误修正预处理等。

(2) 将所述预处理后的文本内容中的每个语句转化为向量序列。

可以对每个语句进行分词或分字。将每个语句的多个词通过训练好的词嵌入模型将每个词转化为词向量；或将每个语句的多个字通过训练好的字嵌入模型将每个字转化为字向量。依词序组合每个语句的词向量得到该语句的向量序列，或依字序组合每个语句的字向量得到该语句的向量序列。

(3) 将每个语句对应的向量序列输入训练好的第一双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第一条件随机场，得到该语句的标记序列。

所述第一双向长短时记忆网络可以提取该语句的语义信息和该语句中每个词（或字）的上下文信息，并以特征向量序列的形式将提取的信息传递给所述第一条件随机场，通过所述第一条件随机场对该语句中的每个词（或字）进行分类，得到该语句的标记序列。

(4) 根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

医疗命名实体可以包括疾病、症状和药品等。例如，医疗文本中的一个语句为“糖尿病是危害人类身心健康的主要慢性疾病之一”的标记序列为“B-1 I-1 E-1 0 0 0 0 0 0 0 0 0 0 0 0”，则该语句的医疗命名实体为疾病“糖尿病”。其中，“B-1”表示疾病（用 1 表示疾病，2 表示症状，3 表示药品）的开始字，“I-1”表示疾病的中间字，“E-1”表示疾病的末尾字，“0”表示其他字（非命名实体的字）。

在另一实施例中，在所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别之前，所述方法还包括：

获取医疗训练语句和所述医疗训练语句的标签向量序列；

将所述医疗训练语句转化为向量序列；

将所述医疗训练语句的向量序列输入所述第一双向长短时记忆网络，得到所述医疗训练语句的特征向量序列，将所述医疗训练语句的特征向量序列输入所述第一条件随机场，得到所述医疗训练语句的标记序列；

根据所述医疗训练语句的标签向量序列和标记序列通过梯度下降法调整所述第一双向长短时记忆网络和所述第一条件随机场的隐藏层的权重和偏置值；

循环执行上述步骤，得到所述训练好的第一双向长短时记忆网络和所述训练好的第一条件随机场。

接收模块 203，用于接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源。

在一具体实施例中，可以通过对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源对所述待搜索文本进行筛选。例如，待搜索文本为“糖尿病治疗”，对待搜索文本的限制时间为“2019”，待搜索文本的类型为“指南”，待搜索文本的目标来源为“中华医学会”。

扩展模块 204，用于对所述待搜索文本进行扩展，得到扩展后的待搜索文本。

在一具体实施例中，所述对所述待搜索文本进行扩展包括：

对所述待搜索文本进行分词处理，得到多个目标词语；

确定所述多个词语中每个目标词语的近义词集合；

在所述待搜索文本中用每个目标词语的近义词集合中的词语替换该目标词语，将每次替换后的所述待搜索文本确定为一个扩展后的待搜索文本。

在另一实施例中，所述确定所述多个词语中每个目标词语的近义词集合包括：

从预设近义词词典中查询该目标词语的近义词；和/或

获取历史搜索信息，从所述历史搜索信息中标记的医疗文本中确定该目标词语的近义词；和/或

从所述待搜索文本中删除该目标词语，得到该目标词语的匹配模板，从所述多个医疗文本中查找所述匹配模板，将查找到的所述匹配模板所在的语句中与该目标词语对应的词语确定为该目标词语的近义词，其中，该目标词语在所述匹配模板中的位置与该目标词语对应的词语在查找到的所述匹配模板中的位置一致。

第二识别模块 205，用于对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体。

在一具体实施例中，所述对所述扩展后的待搜索文本进行医疗命名实体识别包括：

对所述扩展后的待搜索文本进行预处理，得到预处理后的扩展后的待搜索文本；

将所述预处理后的扩展后的待搜索文本中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第二双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第二条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

搜索模块 206，用于在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本。

例如，可以在所述多个医疗文本中搜索医疗命名实体与扩展后的待搜索文本的医疗命名实体（如糖尿病、高血糖等）匹配、来源（如中华医学会）及类型（如指南）与待搜索文本匹配并且发布时间晚于待搜索文本的限制时间（如 2019）的多个候选医疗文本。

计算模块 207，用于根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分。

在一具体实施例中，所述根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分包括：

对于所述多个候选医疗文本中的第 i 个医疗文本， $1 \leq i \leq n$ ， n 为所述多个候选医疗文本的数量，根据第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的位置序号确定第 i 个医疗文本的位置权重 W_i ：

$$W_i = \sum_{j=1}^{m_i} D_{i,j}, \quad D_{i,j} = \begin{cases} d_1, O_{i,j} \in E_{i,1} \\ d_2, O_{i,j} \in E_{i,2} \\ d_3, O_{i,j} \in E_{i,3} \\ d_4, O_{i,j} \in E_{i,4} \end{cases}$$

其中， m_i 为第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的数量， $O_{i,j}$ 表示第 i 个医疗文本中第 j 个医疗命名实体的位置序号， $E_{i,1}$ 表示位于第 i 个医疗文本的标题的位置序号的集合， $E_{i,2}$ 表示位于第 i 个医疗文本的摘要的位置序号的集合， $E_{i,3}$ 表示位于第 i 个医疗文本的正文内容的头尾语句的位置序号的集合， $E_{i,4}$ 表示位于第 i 个医疗文本的正文内容的非头尾语句的位置序号的集合， d_1 、 d_2 、 d_3 和 d_4 为预设值；

根据第 i 个医疗文本的医疗命名实体和所述扩展后的待搜索文本的医疗命名实体计算第 i 个医疗文本的长度匹配度 l_i ：

$$l_i = \frac{C_i}{\max(A_i, B)}$$

其中, C_i 表示第 i 个医疗文本中与所述扩展后的待搜索文本的医疗命名实体一致的医疗命名实体的数量, A_i 表示第 i 个医疗文本中医疗命名实体的数量, B 表示所述扩展后的待搜索文本中医疗命名实体的数量;

从第 i 个医疗文本中查找词语与所述扩展后的待搜索文本的词语一致的目标文本, 计算所述目标文本与所述扩展后的待搜索文本的逆序数 G_i , 根据逆序数 G_i 和所述扩展后的待搜索文本计算第 i 个医疗文本的词序相似度 s_i :

$$s_i = 1 - \frac{G_i}{F - 1}$$

其中, F 表示所述扩展后的待搜索文本中词语的数量;

根据第 i 个医疗文本的位置权重、长度匹配度和词序相似度计算第 i 个医疗文本的匹配得分 P_i :

$$P_i = W_i(w_{i,1}l_i + w_{i,2}s_i)$$

其中, $w_{i,1}$ 和 $w_{i,2}$ 为预设权重。

具体地, $d_1 > d_2 > d_3 > d_4$, d_1 、 d_2 、 d_3 、 d_4 可以分别取值为 4、3、2、1; $w_{i,1}$ 、 $w_{i,2}$ 可以分别取值为 0.5、0.5。

返回模块 208, 用于将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

在一具体实施例中, 可以将匹配得分最高的预设数量的候选医疗文本的摘要和对应的链接地址返回所述请求方。例如, 可以将匹配得分最高的 5 个候选医疗文本的摘要和对应的链接地址返回给请求方, 使请求方可以通过点击该链接地址获取更详细匹配度更高的信息。

实施例二的医疗文本搜索装置 20 获取多个医疗文本, 每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容, 每个医疗文本有对应的链接地址; 对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别, 得到每个医疗文本的多个医疗命名实体, 每个医疗命名实体有对应的位置序号; 接收请求方的搜索请求, 所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源; 对所述待搜索文本进行扩展, 得到扩展后的待搜索文本; 对所述扩展后的待搜索文本进行医疗命名实体识别, 得到所述扩展后的待搜索文本的多个医疗命名实体; 在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本; 根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分; 将匹配得分最高的预设数量的候选医疗文本返回所述请求方。实施例二根据用户的输入从医疗文本集中搜索匹配医疗文本, 提升了医疗文本搜索的准确度。

实施例三

本实施例提供一种可读存储介质, 该可读存储介质上存储有计算机可读指令, 所述可读存储介质可以是非易失性可读存储介质, 也可以是易失性可读存储介质。该计算机可读指令被处理器执行时实现上述医疗文本搜索方法实施例中的步骤, 例如图 1 所示的步骤 101-108。或者, 该计算机可读指令被处理器执行时实现上述装置实施例中各模块的功能, 例如图 2 中的模块 201-208。

实施例四

图 3 为本申请实施例三提供的计算机设备的示意图。所述计算机设备 30 包括存储器 301、处理器 302 以及存储在所述存储器 301 中并可在所述处理器 302 上运行的计算机可读指令 303, 例如医疗文本搜索程序。所述处理器 302 执行所述计算机可读指令 303 时实现上述医疗文本搜索方法实施例中的步骤, 例如图 1 所示的 101-108。或者, 该计算机可读指令被处理器执行时实现上述装置实施例中各模块的功能, 例如图 2 中的模块 201-208。

示例性的, 所述计算机可读指令 303 可以被分割成一个或多个模块, 所述一个或者多个

模块被存储在所述存储器 301 中，并由所述处理器 302 执行，以完成本方法。所述一个或多个模块可以是能够完成特定功能的一系列计算机可读指令的指令段，该指令段用于描述所述计算机可读指令 303 在所述计算机设备 30 中的执行过程。例如，所述计算机可读指令 303 可以被分割成图 2 中的获取模块 201、第一识别模块 202、接收模块 203、扩展模块 204、第二识别模块 205、搜索模块 206、计算模块 207、返回模块 208，各模块具体功能参见实施例二。

本领域技术人员可以理解，所述示意图 3 仅仅是计算机设备 30 的示例，并不构成对计算机设备 30 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述计算机设备 30 还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 302 可以是中央处理单元 (Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器 302 也可以是任何常规的处理器等，所述处理器 302 是所述计算机设备 30 的控制中心，利用各种接口和线路连接整个计算机设备 30 的各个部分。

所述存储器 301 可用于存储所述计算机可读指令 303，所述处理器 302 通过运行或执行存储在所述存储器 301 内的计算机可读指令，以及调用存储在存储器 301 内的数据，实现所述计算机设备 30 的各种功能。所述存储器 301 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如声音播放功能、图像播放功能等）等；存储数据区可存储根据计算机设备 30 的使用所创建的数据（比如音频数据等）等。此外，存储器 301 可以包括非易失性存储器，例如硬盘、内存、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)、至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。

所述计算机设备 30 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，可以存储在一个计算机可读存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，也可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一可读存储介质中，该计算机可读指令在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机可读指令包括计算机可读指令代码，所述计算机可读指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器 (ROM, Read-Only Memory)。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

上述以软件功能模块的形式实现的集成的模块，可以存储在一个计算机可读存储介质中。上述软件功能模块存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）或处理器 (processor) 执行本申请各个实施例所述方法的部分步骤。

对于本领域技术人员而言，显然本申请不限于上述示范性实施例的细节，而且在不背离本申请的精神或基本特征的情况下，能够以其他的具体形式实现本申请。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本申请的范围由所附权利要求

而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。此外，显然“包括”一词不排除其他模块或步骤，单数不排除复数。系统权利要求中陈述的多个模块或装置也可以由一个模块或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

权 利 要 求 书

1. 一种医疗文本搜索方法，其中，所述方法包括：

获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；

接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

2. 如权利要求 1 所述的方法，其中，所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别包括：

对每个医疗文本的标题、摘要、正文内容进行分句、去重、错误修正预处理，得到预处理后的文本内容；

将所述预处理后的文本内容中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第一双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第一条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

3. 如权利要求 2 所述的方法，其中，在所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别之前，所述方法还包括：

获取医疗训练语句和所述医疗训练语句的标签向量序列；

将所述医疗训练语句转化为向量序列；

将所述医疗训练语句的向量序列输入所述第一双向长短时记忆网络，得到所述医疗训练语句的特征向量序列，将所述医疗训练语句的特征向量序列输入所述第一条件随机场，得到所述医疗训练语句的标记序列；

根据所述医疗训练语句的标签向量序列和标记序列通过梯度下降法调整所述第一双向长短时记忆网络和所述第一条件随机场的隐藏层的权重和偏置值。

4. 如权利要求 1 所述的方法，其中，所述对所述待搜索文本进行扩展包括：

对所述待搜索文本进行分词处理，得到多个目标词语；

确定所述多个词语中每个目标词语的近义词集合；

在所述待搜索文本中用每个目标词语的近义词集合中的词语替换该目标词语，将每次替换后的所述待搜索文本确定为一个扩展后的待搜索文本。

5. 如权利要求 4 所述的方法，其中，所述确定所述多个词语中每个目标词语的近义词集合包括：

从预设近义词词典中查询该目标词语的近义词；和/或

获取历史搜索信息，从所述历史搜索信息中标记的医疗文本中确定该目标词语的近义词；和/或

从所述待搜索文本中删除该目标词语，得到该目标词语的匹配模板，从所述多个医疗文本中查找所述匹配模板，将查找到的所述匹配模板所在的语句中与该目标词语对应的词语确定为该目标词语的近义词，其中，该目标词语在所述匹配模板中的位置与该目标词语对应的词语在查找到的所述匹配模板中的位置一致。

6. 如权利要求 1 所述的方法，其中，所述对所述扩展后的待搜索文本进行医疗命名实体识别包括：

对所述扩展后的待搜索文本进行预处理，得到预处理后的扩展后的待搜索文本；

将所述预处理后的扩展后的待搜索文本中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第二双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第二条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

7. 如权利要求 1 所述的方法，其中，所述根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分包括：

对于所述多个候选医疗文本中的第 i 个医疗文本， $1 \leq i \leq n$ ， n 为所述多个候选医疗文本的数量，根据第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的位置序号确定第 i 个医疗文本的位置权重 W_i ：

$$W_i = \sum_{j=1}^{m_i} D_{i,j}, \quad D_{i,j} = \begin{cases} d_1, O_{i,j} \in E_{i,1} \\ d_2, O_{i,j} \in E_{i,2} \\ d_3, O_{i,j} \in E_{i,3} \\ d_4, O_{i,j} \in E_{i,4} \end{cases}$$

其中， m_i 为第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的数量， $O_{i,j}$ 表示第 i 个医疗文本中第 j 个医疗命名实体的位置序号， $E_{i,1}$ 表示位于第 i 个医疗文本的标题的位置序号的集合， $E_{i,2}$ 表示位于第 i 个医疗文本的摘要的位置序号的集合， $E_{i,3}$ 表示位于第 i 个医疗文本的正文内容的头尾语句的位置序号的集合， $E_{i,4}$ 表示位于第 i 个医疗文本的正文内容的非头尾语句的位置序号的集合， d_1 、 d_2 、 d_3 和 d_4 为预设值；

根据第 i 个医疗文本的医疗命名实体和所述扩展后的待搜索文本的医疗命名实体计算第 i 个医疗文本的长度匹配度 l_i ：

$$l_i = \frac{C_i}{\max(A_i, B)}$$

其中， C_i 表示第 i 个医疗文本中与所述扩展后的待搜索文本的医疗命名实体一致的医疗命名实体的数量， A_i 表示第 i 个医疗文本中医疗命名实体的数量， B 表示所述扩展后的待搜索文本中医疗命名实体的数量；

从第 i 个医疗文本中查找词语与所述扩展后的待搜索文本的词语一致的目标文本，计算所述目标文本与所述扩展后的待搜索文本的逆序数 G_i ，根据逆序数 G_i 和所述扩展后的待搜索文本计算第 i 个医疗文本的词序相似度 s_i ： $s_i = 1 - \frac{G_i}{F-1}$

其中， F 表示所述扩展后的待搜索文本中词语的数量；

根据第 i 个医疗文本的位置权重、长度匹配度和词序相似度计算第 i 个医疗文本的匹配得分 P_i ：

$$P_i = W_i(w_{i,1}l_i + w_{i,2}s_i)$$

其中， $w_{i,1}$ 和 $w_{i,2}$ 为预设权重。

8. 一种医疗文本搜索装置，其中，所述装置包括：

获取模块，用于获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

第一识别模块，用于对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的

位置序号；

接收模块，用于接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

扩展模块，用于对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

第二识别模块，用于对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

搜索模块，用于在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

计算模块，用于根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

返回模块，用于将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

9. 如权利要求 8 所述的医疗文本搜索装置，其中，所述第一识别模块具体用于：

对每个医疗文本的标题、摘要、正文内容进行分句、去重、错误修正预处理，得到预处理后的文本内容；

将所述预处理后的文本内容中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第一双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第一条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

10. 如权利要求 9 所述的医疗文本搜索装置，其中，在所述对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别之前，所述第一识别模块具体还用于：

获取医疗训练语句和所述医疗训练语句的标签向量序列；

将所述医疗训练语句转化为向量序列；

将所述医疗训练语句的向量序列输入所述第一双向长短时记忆网络，得到所述医疗训练语句的特征向量序列，将所述医疗训练语句的特征向量序列输入所述第一条件随机场，得到所述医疗训练语句的标记序列；

根据所述医疗训练语句的标签向量序列和标记序列通过梯度下降法调整所述第一双向长短时记忆网络和所述第一条件随机场的隐藏层的权重和偏置值。

11. 如权利要求 8 所述的医疗文本搜索装置，其中，所述扩展模块具体用于：

对所述待搜索文本进行分词处理，得到多个目标词语；

确定所述多个词语中每个目标词语的近义词集合；

在所述待搜索文本中用每个目标词语的近义词集合中的词语替换该目标词语，将每次替换后的所述待搜索文本确定为一个扩展后的待搜索文本。

12. 如权利要求 10 所述的医疗文本搜索装置，其中，所述扩展模块具体还用于：

从预设近义词词典中查询该目标词语的近义词；和/或

获取历史搜索信息，从所述历史搜索信息中标记的医疗文本中确定该目标词语的近义词；和/或

从所述待搜索文本中删除该目标词语，得到该目标词语的匹配模板，从所述多个医疗文本中查找所述匹配模板，将查找到的所述匹配模板所在的语句中与该目标词语对应的词语确定为该目标词语的近义词，其中，该目标词语在所述匹配模板中的位置与该目标词语对应的词语在查找到的所述匹配模板中的位置一致。

13. 如权利要求 8 所述的医疗文本搜索装置，其中，所述第二识别模块具体用于：

对所述扩展后的待搜索文本进行预处理，得到预处理后的扩展后的待搜索文本；

将所述预处理后的扩展后的待搜索文本中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第二双向长短时记忆网络，得到该语句的特征

向量序列, 将该语句的特征向量序列输入训练好的第二条件随机场, 得到该语句的标记序列;
根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

14. 如权利要求 8 所述的医疗文本搜索装置, 其中, 所述计算模块具体用于:

对于所述多个候选医疗文本中的第 i 个医疗文本, $1 \leq i \leq n$, n 为所述多个候选医疗文本的数量, 根据第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的位置序号确定第 i 个医疗文本的位置权重 W_i :

$$W_i = \sum_{j=1}^{m_i} D_{i,j}, \quad D_{i,j} = \begin{cases} d_1, O_{i,j} \in E_{i,1} \\ d_2, O_{i,j} \in E_{i,2} \\ d_3, O_{i,j} \in E_{i,3} \\ d_4, O_{i,j} \in E_{i,4} \end{cases}$$

其中, m_i 为第 i 个医疗文本中与所述扩展后的待搜索文本的多个医疗命名实体一致的医疗命名实体的数量, $O_{i,j}$ 表示第 i 个医疗文本中第 j 个医疗命名实体的位置序号, $E_{i,1}$ 表示位于第 i 个医疗文本的标题的位置序号的集合, $E_{i,2}$ 表示位于第 i 个医疗文本的摘要的位置序号的集合, $E_{i,3}$ 表示位于第 i 个医疗文本的正文内容的头尾语句的位置序号的集合, $E_{i,4}$ 表示位于第 i 个医疗文本的正文内容的非头尾语句的位置序号的集合, d_1 、 d_2 、 d_3 和 d_4 为预设值;

根据第 i 个医疗文本的医疗命名实体和所述扩展后的待搜索文本的医疗命名实体计算第 i 个医疗文本的长度匹配度 l_i :

$$l_i = \frac{C_i}{\max(A_i, B)}$$

其中, C_i 表示第 i 个医疗文本中与所述扩展后的待搜索文本的医疗命名实体一致的医疗命名实体的数量, A_i 表示第 i 个医疗文本中医疗命名实体的数量, B 表示所述扩展后的待搜索文本中医疗命名实体的数量;

从第 i 个医疗文本中查找词语与所述扩展后的待搜索文本的词语一致的目标文本, 计算所述目标文本与所述扩展后的待搜索文本的逆序数 G_i , 根据逆序数 G_i 和所述扩展后的待搜索文本计算第 i 个医疗文本的词序相似度 s_i : $s_i = 1 - \frac{G_i}{F-1}$

其中, F 表示所述扩展后的待搜索文本中词语的数量;

根据第 i 个医疗文本的位置权重、长度匹配度和词序相似度计算第 i 个医疗文本的匹配得分 P_i :

$$P_i = W_i(w_{i,1}l_i + w_{i,2}s_i)$$

其中, $w_{i,1}$ 和 $w_{i,2}$ 为预设权重。

15. 一种计算机设备, 其中, 所述计算机设备包括处理器, 所述处理器用于执行存储器中存储的计算机可读指令以实现如下步骤:

获取多个医疗文本, 每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容, 每个医疗文本有对应的链接地址;

对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别, 得到每个医疗文本的多个医疗命名实体, 每个医疗命名实体有对应的位置序号;

接收请求方的搜索请求, 所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源;

对所述待搜索文本进行扩展, 得到扩展后的待搜索文本;

对所述扩展后的待搜索文本进行医疗命名实体识别, 得到所述扩展后的待搜索文本的多个医疗命名实体;

在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本;

根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分;

将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

16. 一个或多个存储有计算机可读指令的可读存储介质，其中，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如下步骤：

获取多个医疗文本，每个医疗文本包括发布时间、来源、类型、标题、摘要和正文内容，每个医疗文本有对应的链接地址；

对所述多个医疗文本中的每个医疗文本的标题、摘要、正文内容进行医疗命名实体识别，得到每个医疗文本的多个医疗命名实体，每个医疗命名实体有对应的位置序号；

接收请求方的搜索请求，所述搜索请求包括待搜索文本、对所述待搜索文本的限制时间、所述待搜索文本的类型和所述待搜索文本的目标来源；

对所述待搜索文本进行扩展，得到扩展后的待搜索文本；

对所述扩展后的待搜索文本进行医疗命名实体识别，得到所述扩展后的待搜索文本的多个医疗命名实体；

在所述多个医疗文本中搜索医疗命名实体与所述扩展后的待搜索文本的医疗命名实体匹配、来源及类型与所述待搜索文本匹配并且发布时间晚于所述待搜索文本的限制时间的多个候选医疗文本；

根据所述多个候选医疗文本的医疗命名实体的位置序号和所述扩展后的待搜索文本的多个医疗命名实体计算所述多个候选医疗文本的匹配得分；

将匹配得分最高的预设数量的候选医疗文本返回所述请求方。

17. 如权利要求 17 所述的可读存储介质，其中，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

对每个医疗文本的标题、摘要、正文内容进行分句、去重、错误修正预处理，得到预处理后的文本内容；

将所述预处理后的文本内容中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第一双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第一条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

18. 如权利要求 17 所述的可读存储介质，其中，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

对所述待搜索文本进行分词处理，得到多个目标词语；

确定所述多个词语中每个目标词语的近义词集合；

在所述待搜索文本中用每个目标词语的近义词集合中的词语替换该目标词语，将每次替换后的所述待搜索文本确定为一个扩展后的待搜索文本。

19. 如权利要求 17 所述的可读存储介质，其中，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

对所述扩展后的待搜索文本进行预处理，得到预处理后的扩展后的待搜索文本；

将所述预处理后的扩展后的待搜索文本中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第二双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第二条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。

20. 如权利要求 17 所述的可读存储介质，其中，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

对所述扩展后的待搜索文本进行预处理，得到预处理后的扩展后的待搜索文本；

将所述预处理后的扩展后的待搜索文本中的每个语句转化为向量序列；

将每个语句对应的向量序列输入训练好的第二双向长短时记忆网络，得到该语句的特征向量序列，将该语句的特征向量序列输入训练好的第二条件随机场，得到该语句的标记序列；

根据所述预处理后的文本内容中的每个语句的标记序列识别医疗命名实体。



图 1

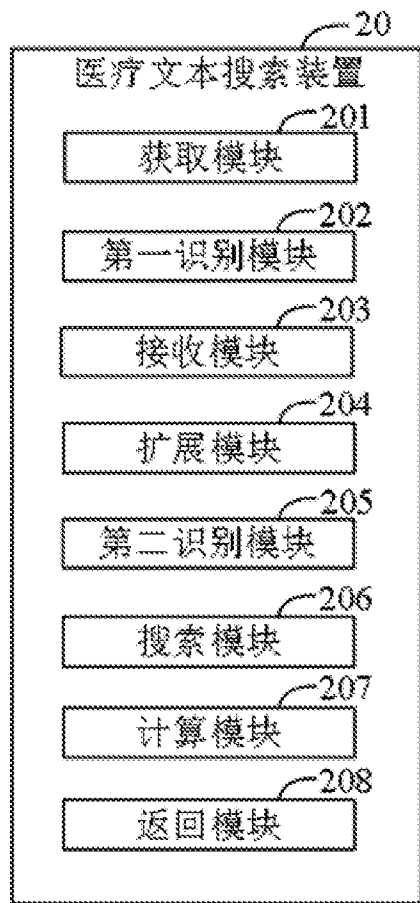


图 2

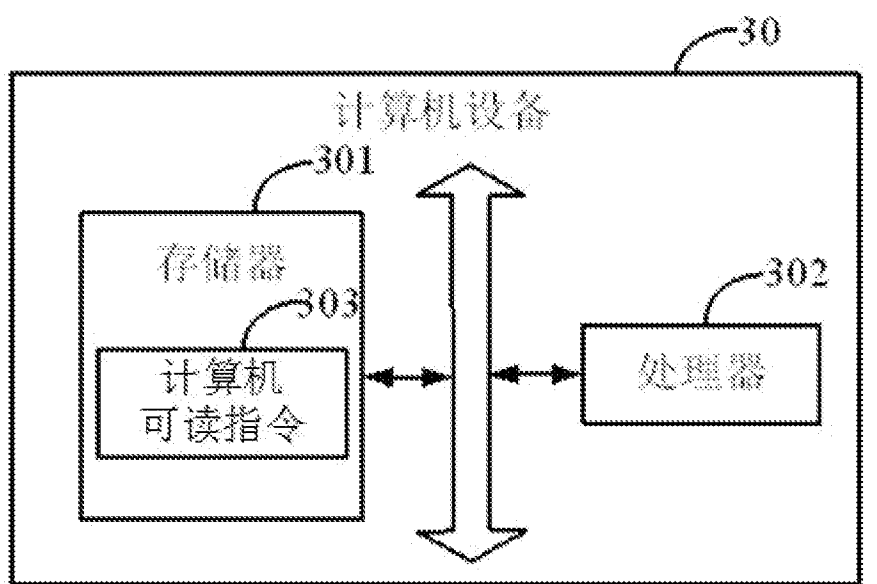


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/093322

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/33(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F;G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; DWPI; SIPOABS; USTXT; EPTXT; WOTXT: 发布, 内容, 信息, 数据, 消息, 文件, 文本, 文字, 资源, 时间, 时长, 时刻, 时段, 间隔, 周期, 日期, 地址, 来源, 类型, 标识, 出处, 位置, 数据源, 源头, 标题, 题目, 正文, 医疗, 诊断, 诊治, 搜索, 检索, 查找, 寻找, 询问, 发现, 命名, 取名, 扩展, 扩充, 同义, 近义, 输入, 输出, 接收, 操作, 设置, 选择, 匹配, 比较, 对比, 比对, 对应, 一致, 相同, 配对, 关联, 验证, 识别, 判断, 辨别, 认证, 辨识, 鉴别, 确定, publish, edition, content, data, log, time, duration, period, date, address, location, source, title, text, body, medical, medicine, health, search, find, name, naming, expand, input, output, write, match, recognition, recognize, identify

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012036160 A1 (SEVENSTER MERLIJN et al.) 09 February 2012 (2012-02-09) description paragraphs [0011]-[0023], [0104]-[0128]	1-6, 8-13, 15-20
Y	CN 109299239 A (FUJIAN HONGYANG SOFT CO., LTD.) 01 February 2019 (2019-02-01) description paragraphs [0070], [0105]	1-6, 8-13, 15-20
Y	CN 106407387 A (SOUTH CHINA NORMAL UNIVERSITY) 15 February 2017 (2017-02-15) description, paragraphs [0019]-[0026]	2-3, 6, 9-10, 12-13, 17-20
Y	US 2008120296 A1 (GENERAL ELECTRIC COMPANY) 22 May 2008 (2008-05-22) description paragraphs [0013]-[0068]	1-6, 8-13, 15-20
Y	US 2016048655 A1 (ACCENTURE GLOBAL SERVICES LTD.) 18 February 2016 (2016-02-18) description, paragraphs [0031]-[0080]	1-6, 8-13, 15-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

17 December 2020

Date of mailing of the international search report

28 December 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/093322

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015302084 A1 (STEWART ROBERT) 22 October 2015 (2015-10-22) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/093322

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2012036160	A1	09 February 2012	JP	5684788	B2	18 March 2015
				RU	2011146630	A	27 May 2013
				US	8935287	B2	13 January 2015
				WO	2010119421	A1	21 October 2010
				BR	PI1006431	A2	02 June 2020
				JP	2012524321	A	11 October 2012
				EP	2419847	A1	22 February 2012
				RU	2544797	C2	20 March 2015
				CN	102395974	A	28 March 2012
				CN	102395974	B	19 August 2015
CN	109299239	A	01 February 2019	None			
CN	106407387	A	15 February 2017	CN	106407387	B	22 November 2019
US	2008120296	A1	22 May 2008	JP	2008130094	A	05 June 2008
				US	8037052	B2	11 October 2011
US	2016048655	A1	18 February 2016	AU	2015213399	A1	03 March 2016
				AU	2019240633	A1	24 October 2019
				EP	2985711	A1	17 February 2016
				AU	2017210493	A1	17 August 2017
				US	10614196	B2	07 April 2020
US	2015302084	A1	22 October 2015	None			

国际检索报告

国际申请号

PCT/CN2020/093322

A. 主题的分类

G06F 16/33 (2019.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F; G06Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNABS; CNTXT; CNKI; DWPI; SIPOABS; USTXT; EPTXT; WOTXT: 发布, 内容, 信息, 数据, 消息, 文件, 文本, 文字, 资源, 时间, 时长, 时刻, 时段, 间隔, 周期, 日期, 地址, 来源, 类型, 标识, 出处, 位置, 数据源, 源头, 标题, 题目, 正文, 医疗, 诊断, 诊治, 搜索, 检索, 查找, 寻找, 询问, 发现, 命名, 取名, 扩展, 扩充, 同义, 近义, 输入, 输出, 接收, 操作, 设置, 选择, 匹配, 比较, 对比, 比对, 对应, 一致, 相同, 配对, 关联, 验证, 识别, 判断, 辨别, 认证, 辨识, 鉴别, 确定, publish, edition, content, data, log, time, duration, period, date, address, location, source, title, text, body, medical, medicine, health, search, find, name, naming, expand, input, output, write, match, recognition, recognize, identify

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	US 2012036160 A1 (SEVENSTER MERLIJN等) 2012年 2月 9日 (2012 - 02 - 09) 说明书第[0011]-[0023]、[0104]-[0128]段	1-6, 8-13, 15-20
Y	CN 109299239 A (福建弘扬软件股份有限公司) 2019年 2月 1日 (2019 - 02 - 01) 说明书第[0070]、[0105]段	1-6, 8-13, 15-20
Y	CN 106407387 A (华南师范大学) 2017年 2月 15日 (2017 - 02 - 15) 说明书第[0019]-[0026]段	2-3, 6, 9-10, 12-13, 17-20
Y	US 2008120296 A1 (GENERAL ELECTRIC COMPANY) 2008年 5月 22日 (2008 - 05 - 22) 说明书第[0013]-[0068]段	1-6, 8-13, 15-20
Y	US 2016048655 A1 (ACCENTURE GLOBAL SERVICES LTD.) 2016年 2月 18日 (2016 - 02 - 18) 说明书第[0031]-[0080]段	1-6, 8-13, 15-20
A	US 2015302084 A1 (STEWART ROBERT) 2015年 10月 22日 (2015 - 10 - 22) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 12月 17日

国际检索报告邮寄日期

2020年 12月 28日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

郭永强

电话号码 86-010-62412084

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/093322

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
US	2012036160	A1	2012年 2月 9日	JP	5684788	B2	2015年 3月 18日
				RU	2011146630	A	2013年 5月 27日
				US	8935287	B2	2015年 1月 13日
				WO	2010119421	A1	2010年 10月 21日
				BR	PI1006431	A2	2020年 6月 2日
				JP	2012524321	A	2012年 10月 11日
				EP	2419847	A1	2012年 2月 22日
				RU	2544797	C2	2015年 3月 20日
				CN	102395974	A	2012年 3月 28日
				CN	102395974	B	2015年 8月 19日
CN	109299239	A	2019年 2月 1日	无			
CN	106407387	A	2017年 2月 15日	CN	106407387	B	2019年 11月 22日
US	2008120296	A1	2008年 5月 22日	JP	2008130094	A	2008年 6月 5日
				US	8037052	B2	2011年 10月 11日
US	2016048655	A1	2016年 2月 18日	AU	2015213399	A1	2016年 3月 3日
				AU	2019240633	A1	2019年 10月 24日
				EP	2985711	A1	2016年 2月 17日
				AU	2017210493	A1	2017年 8月 17日
				US	10614196	B2	2020年 4月 7日
US	2015302084	A1	2015年 10月 22日	无			