

(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
2020 年 7 月 16 日 (16.07.2020)



(10) 国际公布号
WO 2020/143319 A1

(51) 国际专利分类号:
G06F 16/26 (2019.01)

(21) 国际申请号: PCT/CN2019/118160

(22) 国际申请日: 2019 年 11 月 13 日 (13.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910014362.9 2019 年 1 月 8 日 (08.01.2019) 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: 金戈(JIN, Ge); 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)。徐亮(XU, Liang); 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: 北京市立方律师事务所(LIFANG & PARTNERS); 中国北京市东城区香河园街 1 号院信德京汇中心 12 层, Beijing 100028 (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: KNOWLEDGE GRAPH COMPLETION METHOD AND APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 知识图谱补全方法、装置、计算机设备及存储介质

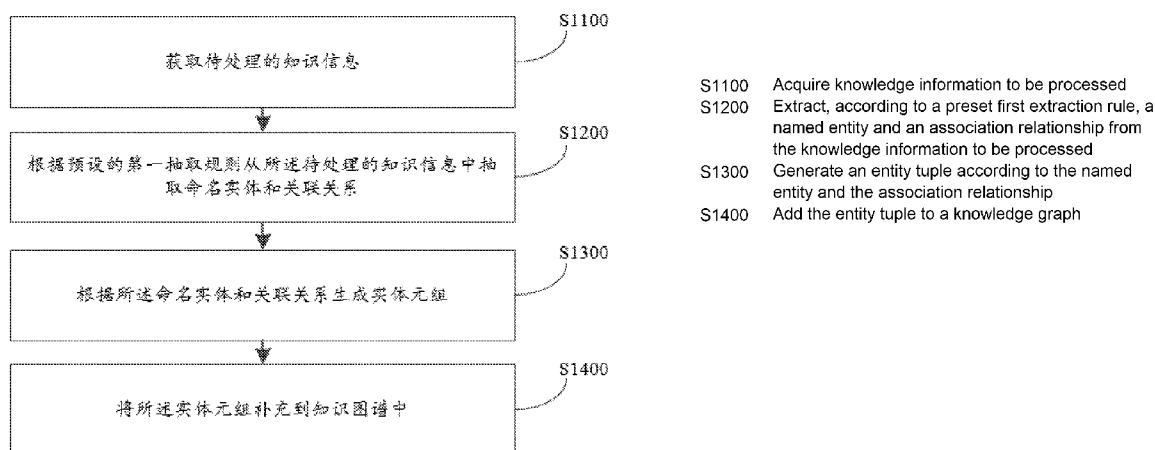


图 1

(57) Abstract: A knowledge graph completion method and apparatus, a computer device and a storage medium. The method comprises the following steps: acquiring knowledge information to be processed (S1100); extracting, according to a preset first extraction rule, a named entity and an association relationship from the knowledge information to be processed (S1200); generating an entity tuple according to the named entity and the association relationship (S1300); and adding the entity tuple to a knowledge graph (S1400). By means of the method, only a small part of acquired data sources needs to be marked, a grammar style is generated according to the marked part of data, and information extraction is then carried out on unmarked data by means of the grammar style, such that the dependency on marking in the data sources can be effectively reduced, the costs of labor and time spent on adding marks to the data sources are reduced, and the efficiency and accuracy of knowledge graph completion are improved.

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,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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)。

根据细则4.17的声明:

— 发明人资格 (细则4.17(iv))

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种知识图谱补全方法、装置、计算机设备及存储介质。所述方法包括下述步骤: 获取待处理的知识信息 (S1100); 根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系 (S1200); 根据所述命名实体和关联关系生成实体元组 (S1300); 将所述实体元组补充到知识图谱中 (S1400)。通过上述方法, 获取到的数据源仅需要少部分标注, 根据有标注的部分数据生成一个文法样式, 再通过该文法样式对未标记的数据进行信息抽取, 可以有效减少对数据源中标注的依赖性, 减少了为数据源添加标注的人力和时间成本, 提高知识图谱补全的效率和精确度。

知识图谱补全方法、装置、计算机设备及存储介质

本申请要求于 2019 年 1 月 8 日提交中国专利局、申请号为 201910014362.9，发明名称为“知识图谱补全方法、装置、计算机设备及存储介质”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

本申请涉及信息处理技术领域，特别是涉及一种知识图谱补全方法、装置、计算机设备及存储介质。

背景技术

知识图谱是知识工程中以图的形式组织的知识集群，其由不同类型的实体作为节点、关系作为连接节点的边所构成的。在知识图谱中，实体指真实世界中的客观物体，或者人类思想中的抽象概念，关系则是描述两个实体之间的实际关系。

在已知的知识图谱中，实体类型有人物、事件、组织机构、地点等，而它们之间的关系类型也十分多样化。不同的实体类型所关注的关系也是不同的。例如，对于人物实体之间，常见关系有亲人及朋友关系；对于人与组织机构之间，常见关系有工作单位、毕业院校等。这些已知的实体间的关系在原始的知识图谱中比较稀疏，而实际上实体间还存在大量的隐含关系。

发明人意识到现有机制中，在构建和补全知识图谱时，会将数据源中所有获取到的实体都进行分析，然后建立所有实体、实体属性之间的关联关系。这样，构建出的知识图谱虽然所能涵盖的范围较广，但对于一批样本，不仅需要知道其中每一个样本数据本身，还需要知道其对应的类别标签，非常依赖含有标签的数据源，标记成本较高。

发明内容

本申请实施例能够提供一种降低对含标注数据源的依赖性的知识图谱补全方法、装置、计算机设备及存储介质。

为解决上述技术问题，本申请实施例提供一种知识图谱补全方法，包括以下步骤：获取待处理的知识信息；根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；根据所述命名实

体和关联关系生成实体元组；将所述实体元组补充到知识图谱中。

为解决上述技术问题，本申请实施例还提供一种知识图谱补全装置，包括：获取模块，用于获取待处理的知识信息；抽取模块，用于根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；处理模块，用于根据所述命名实体和关联关系生成实体元组；执行模块，用于将所述实体元组补充到知识图谱中。

为解决上述技术问题，本申请实施例还提供一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行一种知识图谱补全方法，所述知识图谱补全方法包括以下步骤：获取待处理的知识信息；根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；根据所述命名实体和关联关系生成实体元组；将所述实体元组补充到知识图谱中。

为解决上述技术问题，本申请实施例还提供一种计算机存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行一种知识图谱补全方法，所述知识图谱补全方法包括以下步骤：获取待处理的知识信息；根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；根据所述命名实体和关联关系生成实体元组；将所述实体元组补充到知识图谱中。

本申请实施例通过预设的抽取规则，仅需要对待处理的知识信息中小部分内容进行标注即可实现命名实体和关联关系的抽取，并根据抽取得到的命名实体和关联关系形成实体元组对知识图谱进行补充，相较于传统的需要对待处理数据中所有信息都进行标注才可以实现知识图谱补全的方式，可以有效地减少对于数据源中标注的依赖性，节省对于数据源标注的人力和时间成本，提高了知识图谱补全的效率和准确性。

附图说明

图1为本申请实施例知识图谱补全方法的基本流程示意图；

图2为本申请实施例抽取命名实体和关联关系的流程示意图；

图3为本申请实施例对实体元组进行链接的流程示意图；

图4为本申请实施例对命名实体进行归类并补充关系的流程示意图；

图5为本申请实施例补充属性元组的流程示意图；

图6为本申请实施例补充并列实体元组的流程示意图；

图7为本申请实施例补充从属实体元组的流程示意图；

图8为本申请实施例知识图谱补全装置的基本结构框图；

图9为本申请实施例计算机设备基本结构框图。

具体实施方式

具体地请参阅图 1，图 1 为本实施例知识图谱补全方法的基本流程示意图。如图 1 所示，一种知识图谱补全方法，包括以下步骤：

S1100、获取待处理的知识信息；

在一些实施方式中，可以通过网络爬虫等方式获取到互联网上的内容，包括百科、文章或者文献等等，但不限于此。以百科为例，通过一定的词汇首先获取到该词汇的百科页面，提取其中的文本数据作为目标内容，通过百科中的词汇链接等链接方式，链接到其他的词汇页面中，同时，根据链接方式的不同，确定两个词汇之间的关系，例如同义、反义或者上下位关系等，但不限于此，将获取到的页面内容整理形成文本数据，重复上述方法，统计获取到的大量文本数据作为待处理的知识信息。

S1200、根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；

通过自然语言的解析技术对待处理的知识信息进行命名实体的提取。将待处理的知识信息进行分句，将不同的句子输入到自然语言解析模型中，识别句子中的命名实体，并进行标记，输出得到已标记命名实体的句子，其中，自然语言解析模型可以现有的对自然语言进行分析或识别的任意计算模型。

在待处理的知识信息中确定种子样式，种子样式的确定可以通过人工进行，也可以通过已训练至收敛的用于提取种子样式的神经网络模型进行。种子样式中包含有已标注的句子，标注内容包括句子中的命名实体和关联关系，将一定量的同一类别的种子样式作为训练样本集输入到神经网络模型中，通过反复迭代的方式训练得到收敛的神经网络模型，用于从自然语言句式中提取命名实体和关联关系，定义上述在句子中提取得到命名实体和关联关系的模式为该类别种子样式的文法模式。

将待处理的知识信息的句子根据上述文法模式，识别并抽取命名实体和关联关系，具体地，将上述句子输入到上述用于提取命名实体和关联关系的神经网络模式中，根据神经网络的输出分类结果，得到句子中的命名实体和关联关系。

S1300、根据所述命名实体和关联关系生成实体元组；

根据提取得到的命名实体和关联关系，生成“实体-关系-实体”的三元组，作为实体元组，例如，获取到的实体为“猫”和“哺乳动物”，关系为“种类”，即得生成三元组（猫，种类，哺乳动物）。

S1400、将所述实体元组补充到知识图谱中；

将获取到的实体元组添加到知识图谱中，并判断所述知识图谱中

是否存在与实体元组中命名实体相同的对应实体,若所述知识图谱中已存在与命名实体相同的实体,则将实体元组链接到对应实体;若所述知识图谱中不存在与命名实体相同的对应实体,则根据预设的匹配规则在所述知识图谱中查找所述命名实体的相似实体,并根据相似实体所属的实体元组对所述命名实体进行归类并补充关系。

如图2所示,步骤S1200具体包括以下步骤:

S1210、从所述待处理的知识信息中确定种子样式,其中,所述种子样式包括已标记的语句;

种子样式的确定可以通过人工进行,也可以通过已训练至收敛的用于提取种子样式的神经网络模型进行。种子样式中包含有已标注的语句,标注内容包括语句中的命名实体和关联关系。

S1220、根据所述已标记的语句生成对应的文法模式;

获取到一定量从待处理的知识信息中确定的同一类型的已标记的语句,根据语句的共性确定该类语句的文法模式,其中,文法模式即用于在语句中根据自然语言的表达方法提取出命名实体和关联关系的模式。在一些实施方式中,将一定量的同一类别的种子样式作为训练样本集输入到神经网络模型中,通过反复迭代的方式训练得到收敛的神经网络模型,用于从自然语言句式中提取命名实体和关联关系。

S1230、根据所述文法模式在所述待处理的知识信息中抽取命名实体和关联关系;

将待处理的知识信息进行分句,根据上述方法确定的文法模式,对各个句子进行命名实体和关联关系的提取。在一些实施方式中,可以将分句后的句子输入到已训练至收敛的神经网络模型中,根据神经网络的输出结果确定句子中的命名实体和关联关系。

通过文法模式抽取命名实体和关联关系的方法,只需要在待处理的知识信息中进行一小部分的标记,从而有效降低了对已标记样本的依赖性,节省标记的人力和时间成本。

如图3所示,步骤S1400之后还包括以下步骤:

S1510、判断所述知识图谱中是否存在与所述命名实体相同的对应实体;

将实体元组补充到知识图谱中之后,在知识图谱中查找是否存在与实体元组中的命名实体相同的对应实体,具体地,在获取到实体元组后,将其中的命名实体作为搜索条件,查找知识图谱中搜索与命名实体具有映射关系的实体。

S1520、若所述知识图谱中已存在与所述命名实体相同的对应实体,则将所述实体元组链接到所述对应实体;

当在知识图谱中查找到与命名实体相同的对应实体后,将补充进来的实体元组链接到知识图谱中的对应实体上,作为原有实体的补充

关系，形成对应实体的关系网络。

如图 4 所示，步骤 S1400 之后还包括以下步骤：

S1610、判断所述知识图谱中是否存在与所述命名实体相同的对应实体；

将实体元组补充到知识图谱中之后，在知识图谱中查找是否存在与实体元组中的命名实体相同的对应实体，具体地，在获取到实体元组后，将其中的命名实体作为搜索条件，查找知识图谱中搜索与命名实体具有映射关系的实体。

S1620、若所述知识图谱中不存在与所述命名实体相同的对应实体，则根据预设的匹配规则在所述知识图谱中查找所述命名实体的相似实体；

当知识图谱中查找不到与命名实体相同的对应实体时，根据预设的规则对命名实体进行分类，例如人名、地名或者书名等等，但不限于此，在一些实施方式中，分类的方式可以通过自然语言解析模型对句子的语境进行判断，采用主体识别的方式确定命名实体在句子中所占的语义，并进行分类。在知识图谱中查找分类后的命名实体的相似实体，相似实体与命名实体为同一类别。具体地，查找相似实体的方式可以通过匹配命名实体与知识图谱中同一类别实体的相似度，根据相似度进行排序，确定其中相似度最高的同一类别实体作为命名实体的相似实体。

S1630、根据所述相似实体所属的实体元组对所述命名实体进行归类并补充关系；

提取出相似实体所属的实体元组，根据命名实体与相似实体的匹配关系，对命名实体进行归类并补充关系。具体地，将提取出的相似实体所属的实体元组进行替换，将其中的相似实体替换为命名实体，生成新的实体元组，并根据新生成的实体元组形成命名实体的关系网络，补充知识图谱中的相关内容。例如，获取到命名实体为“田园猫”，查找出的相似实体为“梨花猫”，相似实体所属的实体元组有（梨花猫，种类，哺乳动物），（梨花猫，作用，宠物），将相似实体所属的实体元组中的“梨花猫”替换为“田园猫”，生成（田园猫，种类，哺乳动物），（田园猫，作用，宠物）的实体元组，补充命名实体“田园猫”的关系网络。

如图 5 所示，还包括下述步骤：

S2100、根据预设的第二抽取规则从所述待处理的知识信息中抽取属性元组；

通过预设的第二抽取规则从待处理的知识信息中抽取属性实体、属性名称和属性值，形成三元组，作为属性元组，例如（张三，身高，175CM）。具体地，抽取的方式可以通过已训练至收敛用于在句子中

抽取属性元组的神经网络模型，将待处理的知识信息分句后输入到上述神经网络模型中，根据神经网络的输出结果确定句子中的属性元组，该神经网络模型在训练时，将一定量已标记属性实体、属性名称和属性值的句子作为训练样本集，通过反复迭代的方式直至模型收敛。

S2200、将所述属性元组补充到所述知识图谱中；

将获取到的属性元组添加到知识图谱中，并判断所述知识图谱中是否存在与属性元组中命名实体相同的实体，若所述知识图谱中已存在与属性实体相同的实体，则将属性元组链接到对应实体；若所述知识图谱中不存在与属性实体相同的对应实体，则根据预设的匹配规则在所述知识图谱中查找所述属性实体的相似实体，并根据相似实体所属的属性元组对所述属性实体进行归类并补充关系。

通过上述方法，在实体与关联关系补充的同时，也可以保证实体的属性关系的补充，在抽取命名实体和关联关系后，还可以在知识信息中再次抽取出实体属性和属性值，进一步完善了知识图谱中实体的各项关系和数据生成的关系网络。

如图6所示，步骤S1400之后还包括下述步骤：

S1710、根据预设的计算规则计算所述待处理的知识信息中各命名实体间的并列关系相似度；

将从待处理的知识信息中抽取出的命名实体间进行相似度计算，相似度计算可以依据命名实体所在的句子的相似度，或通过自然语言解析方法对于句子的语义进行判断，并识别出命名实体在句中所充当的成分，得出命名实体的并列关系相似度。另一方面，可以从命名实体所属的实体元组进行并列关系相似度计算，例如实体元组中，关联关系和关联实体相同的命名实体，并列关系相似度较高。

S1720、根据所述并列关系相似度和所述第三抽取规则在所述待处理的知识信息中抽取并列实体元组；

在已有的实体元组中查找出与命名实体并列关系相似度较高的并列实体，根据抽取并列实体的语句模式，在待处理的知识信息中抽取并列实体元组，并确定并列实体的上下位关系。抽取并列实体的语句模式可以是并列实体三元体在所处的句子中抽取出来的语法特征，例如命名实体为“狗”，查找出的具有并列关系的并列实体为“猫”，并且并列实体抽取出的实体元组为(猫，物种，哺乳动物)，即根据该抽取方法，可以抽取出(狗，物种，哺乳动物)作为并列实体元组，表达猫和狗作为并列关系，同属于哺乳动物的下位关系。

S1730、将所述并列实体元组补充到所述知识图谱中。

将抽取到的并列实体元组补充到知识图谱中，并在知识图谱中查找并列实体元组中命名实体相同的实体，若知识图谱中存在与上述命名实体相同的实体，则将并列实体元组链接到与命名实体相同的实体

上,作为实体关系网络的补充,若知识图谱中不存在与命名实体相同的实体,则将该命名实体在知识图谱中作为新的实体。

通过上述并列实体元组的补充方法,可以在补充实体元组之后通过计算得到命名实体的并列和上下位关系,作为原有实体的关系补充,提高了知识图谱中实体间关系网络的完整度。

如图7所示,步骤S1400之后还包括下述步骤:

S1810、根据预设的推理规则判断所述命名实体与所述知识图谱中已存在的实体间的从属关系;

在将实体元组补充到知识图谱之后,在知识图谱中查找与实体元组中的命名实体相关联的已有实体,并且根据推理规则判断命名实体与已有实体间的从属关系。具体地,以命名实体作为搜索条件,在知识图谱中查找与命名实体相同或相似的已有实体,并提取出已有实体所属的实体元组,根据实体元组的关系,将命名实体与另一个实体建立关联关系。例如,补充的实体元组为(张三,同事,李四),在知识图谱中查找“张三”或“李四”,得到已有的实体元组(李四,公司,A企业),即可以逻辑推理的方式获得“张三的公司是A企业”的信息,另一方面,补充的实体元组为(老虎,科,猫科),已有的实体元组为(猫科,目,食肉目),可以推理得到“老虎属于食肉目”的信息。

S1820、根据所述从属关系生成从属实体元组;

根据获取到的从属关系,确定其中的命名实体,生成从属实体元组。例如,根据(老虎,科,猫科),已有的实体元组为(猫科,目,食肉目),可以根据推理结果生成(老虎,目,食肉目)作为从属实体元组。

S1830、将所述从属实体元组补充到所述知识图谱中。

将抽取到的从属实体元组补充到知识图谱中,并在知识图谱中查找从属实体元组中命名实体相同的实体,若知识图谱中存在与上述命名实体相同的实体,则将从属实体元组链接到与命名实体相同的实体上,作为实体关系网络的补充,若知识图谱中不存在与命名实体相同的实体,则将该命名实体在知识图谱中作为新的实体。

通过上述从属实体元组的补充方法,可以根据补充的实体元组和已有的实体元组推理得到新的从属关系,作为原有实体的关系补充,提高了知识图谱中实体间关系网络的完整度。

为解决上述技术问题,本申请实施例还提供一种知识图谱补全装置。请参阅图8,图8为本实施知识图谱补全装置的基本结构框图。

如图8所示,知识图谱补全装置,包括:获取模块2100、抽取模块2200、处理模块2300和执行模块2400。其中,获取模块用于获取待处理的知识信息;抽取模块用于根据预设的第一抽取规则从所述待

处理的知识信息中抽取命名实体和关联关系；处理模块用于根据所述命名实体和关联关系生成实体元组；执行模块用于将所述实体元组补充到知识图谱中。

通过预设的抽取规则，仅需要对待处理的知识信息中小部分内容进行标注即可实现命名实体和关联关系的抽取，并根据抽取得到的命名实体和关联关系形成实体元组对知识图谱进行补充，相较于传统的需要对待处理数据中所有信息都进行标注才可以实现知识图谱补全的方式，可以有效地减少对于数据源中标注的依赖性，节省对于数据源标注的人力和时间成本，提高了知识图谱补全的效率和准确性。

在一些实施方式中，知识图谱补全装置还包括：第一处理子模块、第二处理子模块、第一抽取子模块。其中第一处理子模块用于从所述待处理的知识信息中确定种子样式，其中，所述种子样式包括已标记的语句；第二处理子模块用于根据所述已标记的语句生成对应的文法模式；第一抽取子模块用于根据所述文法模式在所述待处理的知识信息中抽取命名实体和关联关系。

在一些实施方式中，知识图谱补全装置还包括：第一判断子模块、第一执行子模块。其中，第一判断子模块用于判断所述知识图谱中是否存在与所述命名实体相同的对应实体；第一执行子模块用于若所述知识图谱中已存在与所述命名实体相同的对应实体，则将所述实体元组链接到所述对应实体。

在一些实施方式中，知识图谱补全装置还包括：第二判断子模块、第一查找子模块、第二执行子模块。其中，第二判断子模块用于判断所述知识图谱中是否存在与所述命名实体相同的对应实体；第一查找子模块用于若所述知识图谱中不存在与所述命名实体相同的对应实体，则根据预设的匹配规则在所述知识图谱中查找所述命名实体的相似实体；第二执行子模块用于根据所述相似实体所属的实体元组对所述命名实体进行归类并补充关系。

在一些实施方式中，知识图谱补全装置还包括：第二抽取子模块、第三执行子模块。其中，第二抽取子模块用于根据预设的第二抽取规则从所述待处理的知识信息中抽取属性元组；第三执行子模块用于将所述属性元组补充到所述知识图谱中。

在一些实施方式中，知识图谱补全装置还包括：第三处理子模块、第三抽取子模块、第四执行子模块。其中，第三处理子模块用于根据预设的计算规则计算所述待处理的知识信息中各命名实体间的并列关系相似度；第三抽取子模块用于根据所述并列关系相似度和所述第三抽取规则在所述待处理的知识信息中抽取并列实体元组；第四执行子模块用于将所述并列实体元组补充到所述知识图谱中。

在一些实施方式中，知识图谱补全装置还包括：第四处理子模块、

第五处理子模块、第五执行子模块。其中，第四处理子模块用于根据预设的推理规则判断所述命名实体与所述知识图谱中已存在的实体间的从属关系；第五处理子模块用于根据所述从属关系生成从属实体元组；第五执行子模块用于将所述从属实体元组补充到所述知识图谱中。

为解决上述技术问题，本申请实施例还提供一种计算机设备，包括存储器和处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行如上任一项所述的图谱补全方法的步骤。具体请参阅图 9，图 9 为本实施例计算机设备基本结构框图。

如图 9 所示，计算机设备的内部结构示意图。如图 9 所示，该计算机设备包括通过系统总线连接的处理器、非易失性存储介质、存储器和网络接口。其中，该计算机设备的非易失性存储介质存储有操作系统、数据库和计算机可读指令，数据库中可存储有控件信息序列，该计算机可读指令被处理器执行时，可使得处理器实现一种知识图谱补全方法。该计算机设备的处理器用于提供计算和控制能力，支撑整个计算机设备的运行。该计算机设备的存储器中可存储有计算机可读指令，该计算机可读指令被处理器执行时，可使得处理器执行一种知识图谱补全方法。该计算机设备的网络接口用于与终端连接通信。本领域技术人员可以理解，图中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

本实施方式中处理器用于执行图 8 中获取模块 2100、抽取模块 2200、处理模块 2300 和执行模块 2400 的具体功能，存储器存储有执行上述模块所需的程序代码和各类数据。网络接口用于向用户终端或服务器之间的数据传输。本实施方式中的存储器存储有知识图谱补全装置中执行所有子模块所需的程序代码及数据，服务器能够调用服务器的程序代码及数据执行所有子模块的功能。

本申请还提供一种存储有计算机可读指令的存储介质，所述计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行上述任一实施例所述知识图谱补全方法的步骤，其中，存储介质可以为非易失性的。

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。

权 利 要 求 书

1. 一种知识图谱补全方法，包括以下步骤：

获取待处理的知识信息；

根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；

根据所述命名实体和关联关系生成实体元组；

将所述实体元组补充到知识图谱中。

2. 如权利要求 1 所述的知识图谱补全方法，所述根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系的步骤，包括以下步骤：

从所述待处理的知识信息中确定种子样式，其中，所述种子样式包括已标记的语句；

根据所述已标记的语句生成对应的文法模式；

根据所述文法模式在所述待处理的知识信息中抽取命名实体和关联关系。

3. 如权利要求 1 所述的知识图谱补全方法，所述将所述实体元组补充到知识图谱中的步骤之后，包括以下步骤：

判断所述知识图谱中是否存在与所述命名实体相同的对应实体；

若所述知识图谱中已存在与所述命名实体相同的对应实体，则将所述实体元组链接到所述对应实体。

4. 如权利要求 1 所述的知识图谱补全方法，所述将所述实体元组补充到知识图谱中的步骤之后，包括以下步骤：

判断所述知识图谱中是否存在与所述命名实体相同的对应实体；

若所述知识图谱中不存在与所述命名实体相同的对应实体，则根据预设的匹配规则在所述知识图谱中查找所述命名实体的相似实体；

根据所述相似实体所属的实体元组对所述命名实体进行归类并补充关系。

5. 如权利要求 1 所述的知识图谱补全方法，包括下述步骤：

根据预设的第二抽取规则从所述待处理的知识信息中抽取属性元组；

将所述属性元组补充到所述知识图谱中。

6. 如权利要求 1-5 任一项所述的知识图谱补全方法，所述将所述实体元组补充到知识图谱中的步骤之后，包括下述步骤：

根据预设的计算规则计算所述待处理的知识信息中各命名实体间的并列关系相似度；

根据所述并列关系相似度和第三抽取规则在所述待处理的知识信息中抽取并列实体元组；

将所述并列实体元组补充到所述知识图谱中。

7. 如权利要求 1-5 任一项所述的知识图谱补全方法, 所述将所述实体元组补充到知识图谱中的步骤之后, 包括下述步骤:

根据预设的推理规则判断所述命名实体与所述知识图谱中已存在的实体间的从属关系;

根据所述从属关系生成从属实体元组;

将所述从属实体元组补充到所述知识图谱中。

8. 一种知识图谱补全装置, 包括:

获取模块, 用于获取待处理的知识信息;

抽取模块, 用于根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系

处理模块, 用于根据所述命名实体和关联关系生成实体元组;

执行模块, 用于将所述实体元组补充到知识图谱中。

9. 一种计算机设备, 包括:

处理器;

用于存储处理器可执行指令的存储器;

其中, 所述计算机可执行指令被所述处理器执行时, 使得所述处理器执行一种知识图谱补全方法, 所述知识图谱补全方法包括以下步骤:

获取待处理的知识信息;

根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系;

根据所述命名实体和关联关系生成实体元组;

将所述实体元组补充到知识图谱中。

10. 如权利要求 9 所述的计算机设备, 所述根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系的步骤, 包括以下步骤:

从所述待处理的知识信息中确定种子样式, 其中, 所述种子样式包括已标记的语句;

根据所述已标记的语句生成对应的文法模式;

根据所述文法模式在所述待处理的知识信息中抽取命名实体和关联关系。

11. 如权利要求 9 所述的计算机设备, 所述将所述实体元组补充到知识图谱中的步骤之后, 包括以下步骤:

判断所述知识图谱中是否存在与所述命名实体相同的对应实体;

若所述知识图谱中已存在与所述命名实体相同的对应实体, 则将所述实体元组链接到所述对应实体。

12. 如权利要求 9 所述的计算机设备, 所述将所述实体元组补充

到知识图谱中的步骤之后，包括以下步骤：

判断所述知识图谱中是否存在与所述命名实体相同的对应实体；

若所述知识图谱中不存在与所述命名实体相同的对应实体，则根据预设的匹配规则在所述知识图谱中查找所述命名实体的相似实体；

根据所述相似实体所属的实体元组对所述命名实体进行归类并补充关系。

13. 如权利要求 9 所述的计算机设备，包括下述步骤：

根据预设的第二抽取规则从所述待处理的知识信息中抽取属性元组；

将所述属性元组补充到所述知识图谱中。

14. 一种计算机存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行一种知识图谱补全方法，所述知识图谱补全方法包括以下步骤：

获取待处理的知识信息；

根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系；

根据所述命名实体和关联关系生成实体元组；

将所述实体元组补充到知识图谱中。

15. 如权利要求 14 所述的非易失性存储介质，所述根据预设的第一抽取规则从所述待处理的知识信息中抽取命名实体和关联关系的步骤，包括以下步骤：

从所述待处理的知识信息中确定种子样式，其中，所述种子样式包括已标记的语句；

根据所述已标记的语句生成对应的文法模式；

根据所述文法模式在所述待处理的知识信息中抽取命名实体和关联关系。

16. 如权利要求 14 所述的非易失性存储介质，所述将所述实体元组补充到知识图谱中的步骤之后，包括以下步骤：

判断所述知识图谱中是否存在与所述命名实体相同的对应实体；

若所述知识图谱中已存在与所述命名实体相同的对应实体，则将所述实体元组链接到所述对应实体。

17. 如权利要求 14 所述的非易失性存储介质，所述将所述实体元组补充到知识图谱中的步骤之后，包括以下步骤：

判断所述知识图谱中是否存在与所述命名实体相同的对应实体；

若所述知识图谱中不存在与所述命名实体相同的对应实体，则根据预设的匹配规则在所述知识图谱中查找所述命名实体的相似实体；

根据所述相似实体所属的实体元组对所述命名实体进行归类并补充关系。

18. 如权利要求 14 所述的非易失性存储介质，包括下述步骤：

根据预设的第二抽取规则从所述待处理的知识信息中抽取属性元组；

将所述属性元组补充到所述知识图谱中。

19. 如权利要求 14-18 任一项所述的非易失性存储介质，所述将所述实体元组补充到知识图谱中的步骤之后，包括下述步骤：

根据预设的计算规则计算所述待处理的知识信息中各命名实体间的并列关系相似度；

根据所述并列关系相似度和第三抽取规则在所述待处理的知识信息中抽取并列实体元组；

将所述并列实体元组补充到所述知识图谱中。

20. 如权利要求 14-18 任一项所述的非易失性存储介质，所述将所述实体元组补充到知识图谱中的步骤之后，包括下述步骤：

根据预设的推理规则判断所述命名实体与所述知识图谱中已存在的实体间的从属关系；

根据所述从属关系生成从属实体元组；

将所述从属实体元组补充到所述知识图谱中。

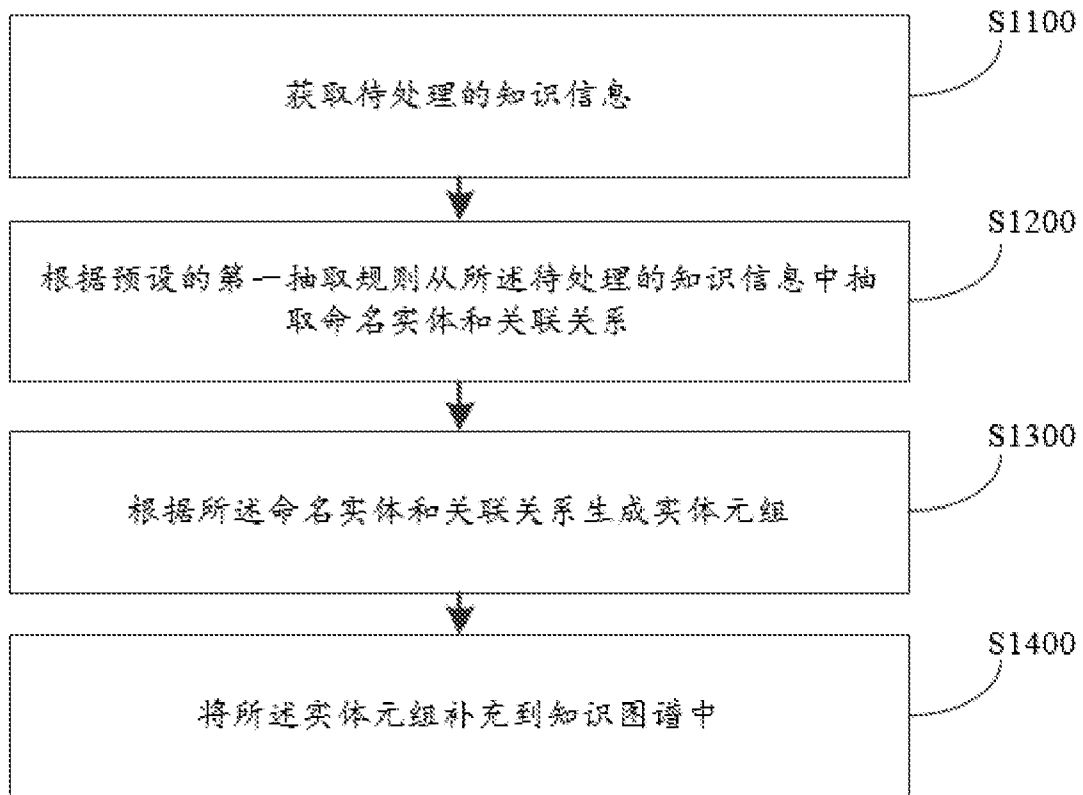


图 1

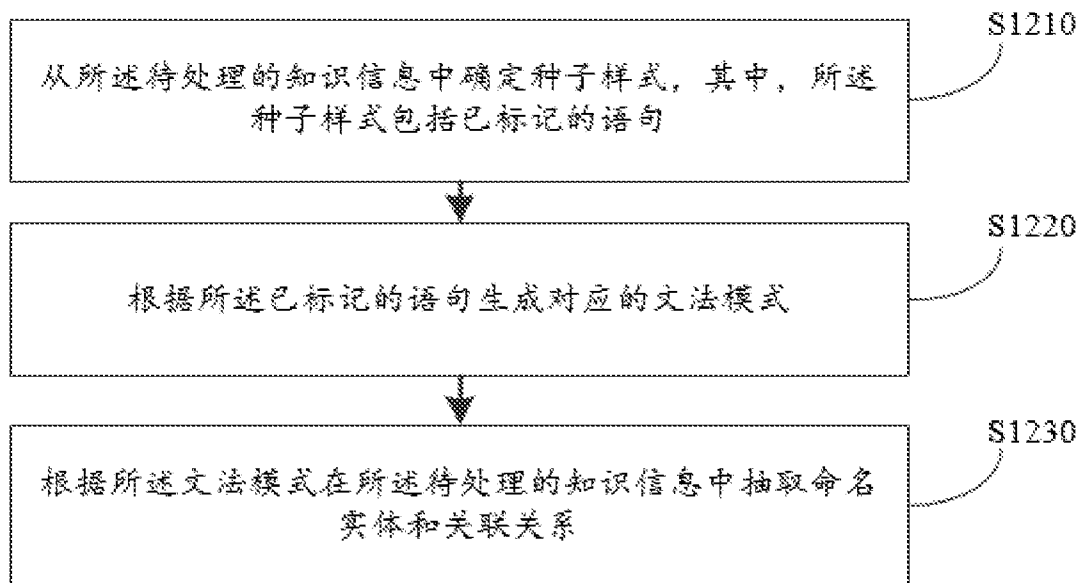


图 2

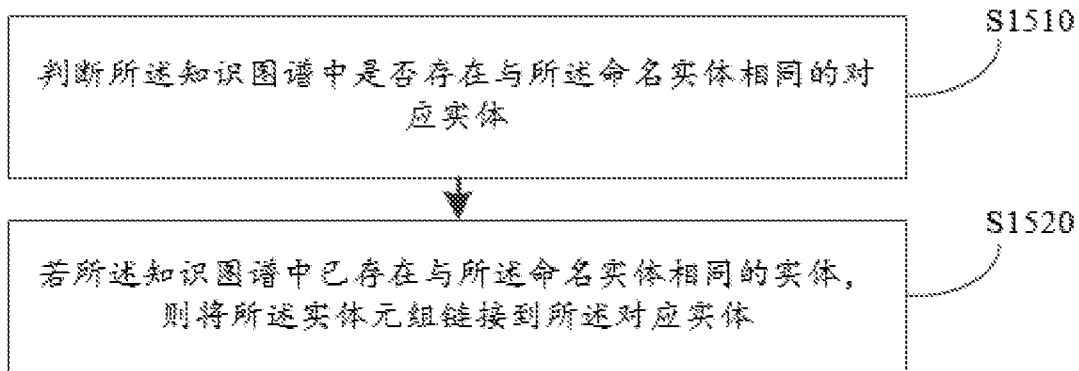


图 3

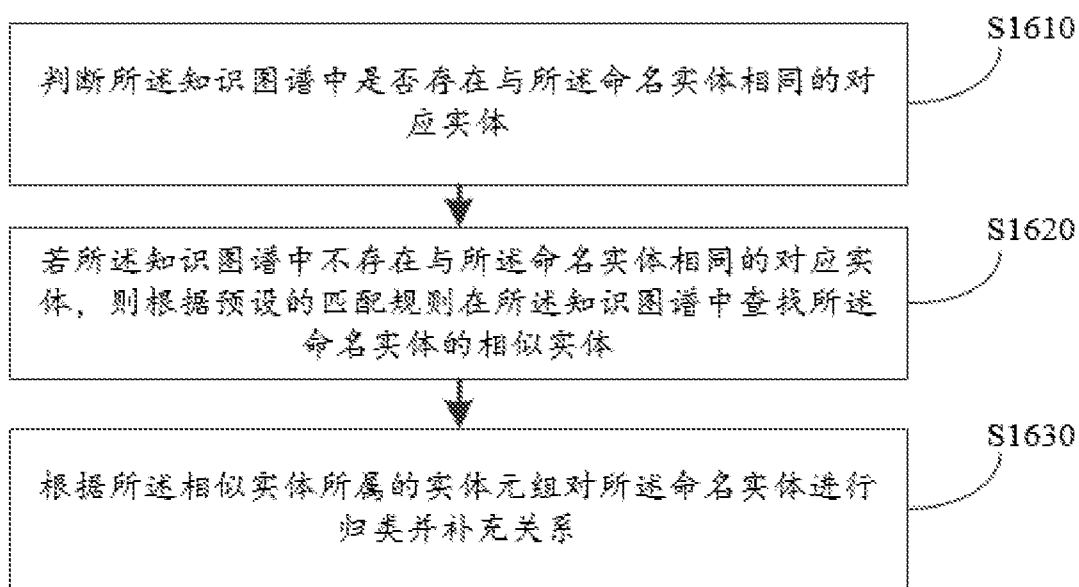


图 4

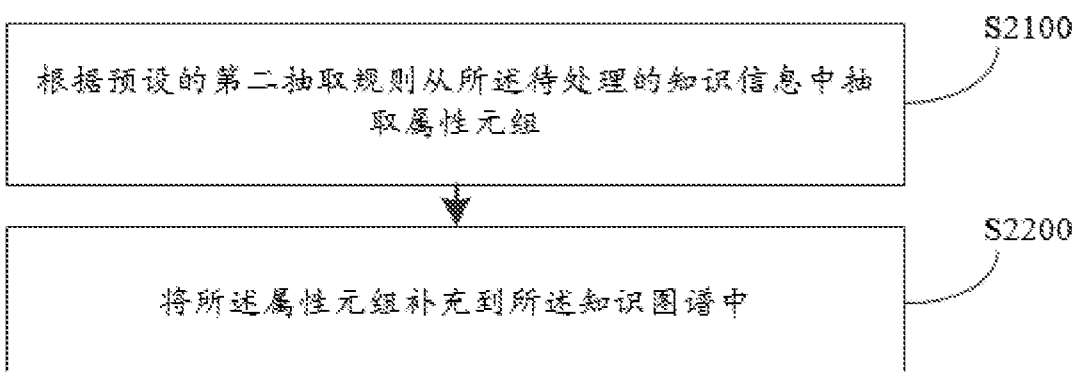


图 5

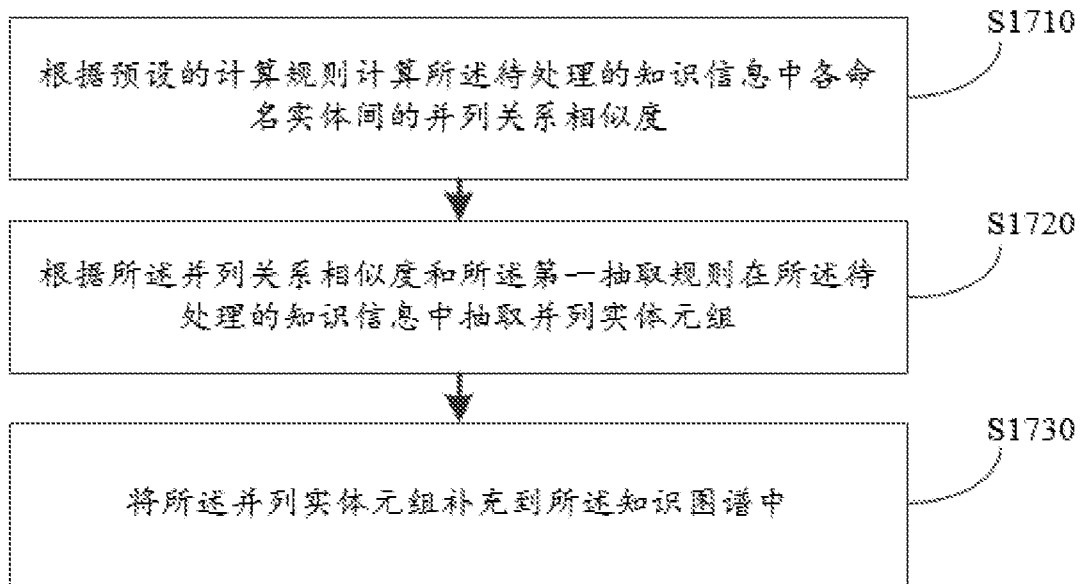


图 6

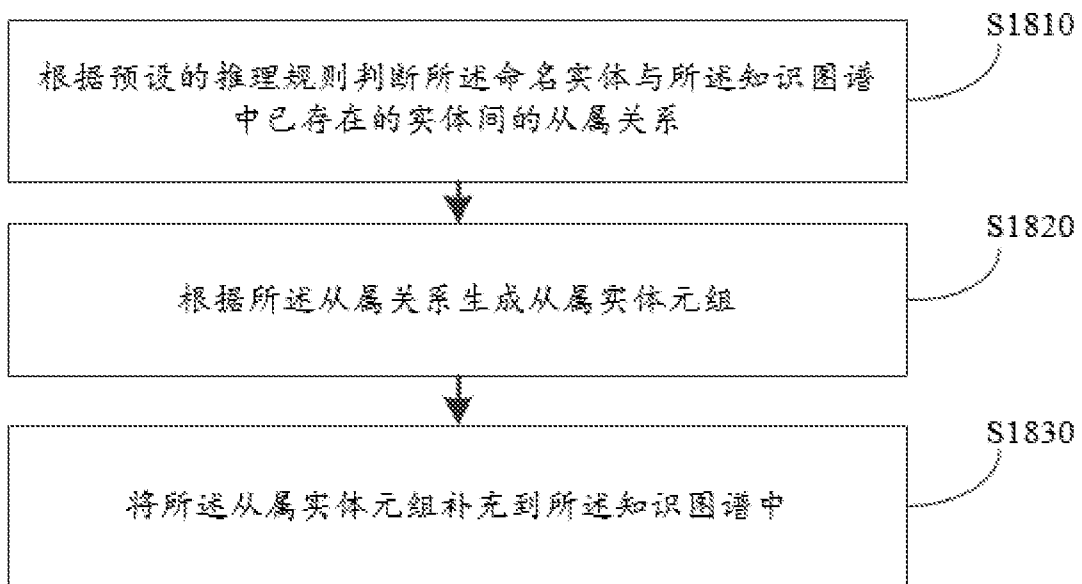


图 7

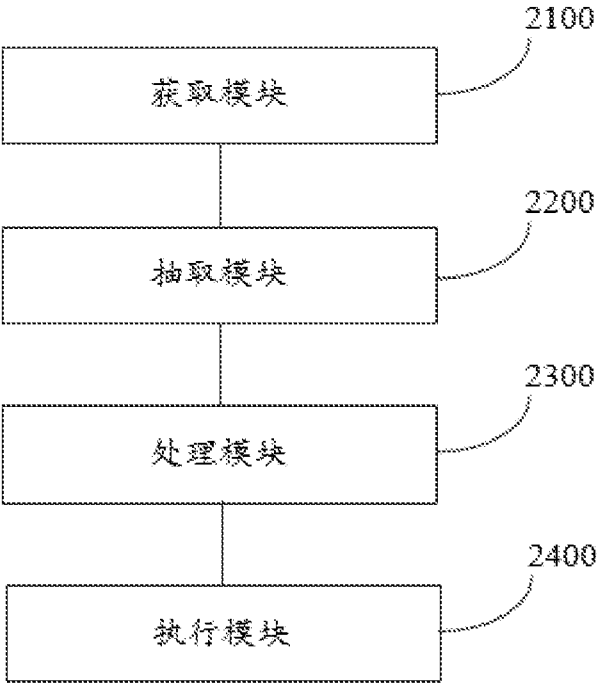


图 8

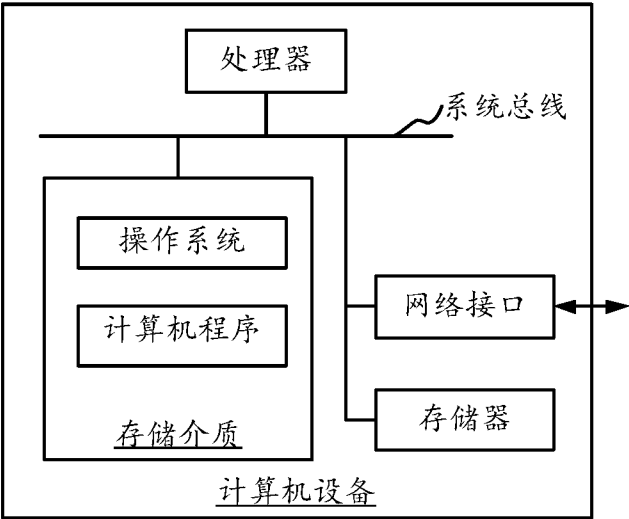


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/118160

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/26(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

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)

CNPAT, WPI, EPODOC, CNKI, IEEE: 知识, 图谱, 实体, 种子, 标注, 标签, 属性, 关系, 关联, 相似, 相近, 近义, 同义, 并列, knowledge, graph, entity, seed, tag, mark, attribute, associat+, relationship, similar, parallel

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107358315 A (GIONEE COMMUNICATION EQUIPMENT CO., LTD.) 17 November 2017 (2017-11-17) description, paragraphs 36-145 and 180-193	1-20
X	CN 104933164 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 23 September 2015 (2015-09-23) description, paragraphs 11-116	1, 8, 9, 14
X	CN 106776711 A (ZHEJIANG UNIVERSITY) 31 May 2017 (2017-05-31) description, paragraphs 74-140	1, 8, 9, 14
X	CN 103488724 A (FUDAN UNIVERSITY) 01 January 2014 (2014-01-01) description, paragraphs 6-48	1, 8, 9, 14
PX	CN 109885691 A (PING'AN TECHNOLOGY (SHENZHEN) CO., LTD.) 14 June 2019 (2019-06-14) claims 1-10	1-20



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

19 January 2020

Date of mailing of the international search report

27 February 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
Information on patent family members

International application No.

PCT/CN2019/118160

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107358315	A	17 November 2017	None			
CN	104933164	A	23 September 2015	CN	104933164	B	09 October 2018
CN	106776711	A	31 May 2017	None			
CN	103488724	A	01 January 2014	CN	103488724	B	28 September 2016
CN	109885691	A	14 June 2019	None			

国际检索报告

国际申请号

PCT/CN2019/118160

A. 主题的分类 G06F 16/26 (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 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)) CNPAT, WPI, EPDOC, CNKI, IEEE: 知识, 图谱, 实体, 种子, 标注, 标签, 属性, 关系, 关联, 相似, 相近, 近义, 同义, 并列, knowledge, graph, entity, seed, tag, mark, attribute, associat+, relationship, similar, parallel																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 107358315 A (深圳市金立通信设备有限公司) 2017年 11月 17日 (2017 - 11 - 17) 说明书第36-145、180-193段</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 104933164 A (华南理工大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第11-116段</td> <td>1、8-9、14</td> </tr> <tr> <td>X</td> <td>CN 106776711 A (浙江大学) 2017年 5月 31日 (2017 - 05 - 31) 说明书第74-140段</td> <td>1、8-9、14</td> </tr> <tr> <td>X</td> <td>CN 103488724 A (复旦大学) 2014年 1月 1日 (2014 - 01 - 01) 说明书第6-48段</td> <td>1、8-9、14</td> </tr> <tr> <td>PX</td> <td>CN 109885691 A (平安科技深圳有限公司) 2019年 6月 14日 (2019 - 06 - 14) 权利要求1-10</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 107358315 A (深圳市金立通信设备有限公司) 2017年 11月 17日 (2017 - 11 - 17) 说明书第36-145、180-193段	1-20	X	CN 104933164 A (华南理工大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第11-116段	1、8-9、14	X	CN 106776711 A (浙江大学) 2017年 5月 31日 (2017 - 05 - 31) 说明书第74-140段	1、8-9、14	X	CN 103488724 A (复旦大学) 2014年 1月 1日 (2014 - 01 - 01) 说明书第6-48段	1、8-9、14	PX	CN 109885691 A (平安科技深圳有限公司) 2019年 6月 14日 (2019 - 06 - 14) 权利要求1-10	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 107358315 A (深圳市金立通信设备有限公司) 2017年 11月 17日 (2017 - 11 - 17) 说明书第36-145、180-193段	1-20																		
X	CN 104933164 A (华南理工大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第11-116段	1、8-9、14																		
X	CN 106776711 A (浙江大学) 2017年 5月 31日 (2017 - 05 - 31) 说明书第74-140段	1、8-9、14																		
X	CN 103488724 A (复旦大学) 2014年 1月 1日 (2014 - 01 - 01) 说明书第6-48段	1、8-9、14																		
PX	CN 109885691 A (平安科技深圳有限公司) 2019年 6月 14日 (2019 - 06 - 14) 权利要求1-10	1-20																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 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 </td> <td> “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 </td> </tr> </table>			* 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																
* 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 1月 19日		国际检索报告邮寄日期 2020年 2月 27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		授权官员 武晓冬 电话号码 86-(10)-53961385																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/118160

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)
CN	107358315	A	2017年 11月 17日	无	
CN	104933164	A	2015年 9月 23日	CN 104933164	B 2018年 10月 9日
CN	106776711	A	2017年 5月 31日	无	
CN	103488724	A	2014年 1月 1日	CN 103488724	B 2016年 9月 28日
CN	109885691	A	2019年 6月 14日	无	