

## 知识图谱数据融合

### 技术领域

[01]本申请涉及数据处理技术领域，特别涉及知识图谱数据融合方法和系统。

### 背景技术

[02]不同平台或不同业务领域分别拥有各自的数据。随着数据管理和数据建设的发展，希望能够对多平台、多业务领域的数据进行融合和连通。知识图谱是一种结构化的数据表达方式，能够高效的呈现数据所蕴含的知识信息。如果通过知识图谱实现多平台、多业务领域的知识连通将可以有效提升数据融合的效率，带来业务效果以及计算效能的提升。

[03]因此，亟需知识图谱数据融合方法和系统，来实现数据的融合和连通。

### 发明内容

[04]本说明书一个方面提供一种知识图谱数据融合方法，包括：获取目标实体字段以及目标关系描述；所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据；其中，知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述；确定一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子；从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱。

[05]本说明书另一个方面提供一种知识图谱数据融合系统，包括：目标数据获取模块，用于获取目标实体字段以及目标关系描述；所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据；其中，知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述；图谱算子确定模块，用于确定一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子；融合图谱生成模块，用于从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱。

[06]本说明书另一个方面提供一种知识图谱数据融合装置，包括至少一个存储介质和至少一个处理器，所述至少一个存储介质用于存储计算机指令；所述至少一个处理器用于执行所述计算机指令以实现所述的知识图谱数据融合方法。

[07]本说明书一个方面提供一种知识图谱数据处理方法，包括：向服务方指定目标实体字段以及目标关系描述；所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据；其中，知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述；从所述服务方处获取融合知识图谱和/或从所述服务方处获取目标任务结果；所述融合知识图谱通过图谱算子处理数据实例生成，所述数据实例基于所述目标实体字段以及所述目标关系描述从所述两个或更多个知识图谱中获取；所述目标任务结果通过目标任务算法处理所述融合知识图谱得到；所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

[08]本说明书另一个方面提供一种知识图谱数据处理系统，包括：目标数据指定模块，用于向服务方指定目标实体字段以及目标关系描述；所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据；其中，知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述；结果获取模块，用于从所述服务方处获取融合知识图谱和/或从所述服务方处获取目标任务结果；所述融合知识图谱通过图谱算子处理数据实例生成，所述数据实例基于所述目标实体字段以及所述目标关系描述从所述两个或更多个知识图谱中获取；所述目标任务结果通过目标任务算法处理所述融合知识图谱得到；所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

[09]本说明书另一个方面提供一种知识图谱数据处理装置，包括至少一个存储介质和至少一个处理器，所述至少一个存储介质用于存储计算机指令；所述至少一个处理器用于执行所述计算机指令以实现所述的知识图谱数据处理方法。

### 附图说明

[10]本说明书将以示例性实施例的方式进一步说明，这些示例性实施例将通过附图进行详细描述。这些实施例并非限制性的，在这些实施例中，相同的编号表示相同的结构，其中：

[11]图 1 是根据本说明书一些实施例所示的一种知识图谱数据融合系统的应用场景示意图；

[12]图 2 是根据本说明书一些实施例所示的一种知识图谱数据融合系统的框图；

[13]图 3 是根据本说明书一些实施例所示的一种知识图谱数据融合方法的示例性流程图；

[14]图 4 是根据本说明书一些实施例所示的一种融合知识图谱的本体定义数据的示意图；

[15]图 5 是根据本说明书一些实施例所示的一种生成融合知识图谱的示例性流程图；

[16]图 6 是根据本说明书一些实施例所示的一种知识图谱数据处理方法的示例性流程图。

### 具体实施方式

[17]为了更清楚地说明本说明书实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单的介绍。显而易见地，下面描述中的附图仅仅是本说明书的一些示例或实施例，对于本领域的普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图将本说明书应用于其它类似情景。除非从语言环境中显而易见或另做说明，图中相同标号代表相同结构或操作。

[18]应当理解，本说明书中所使用的“系统”、“装置”、“单元”和/或“模组”是用于区分不同级别的不同组件、元件、部件、部分或装配的一种方法。然而，如果其他词语可实现相同的目的，则可通过其他表达来替换所述词语。

[19]如本说明书和权利要求书所示，除非上下文明确提示例外情形，“一”、“一个”、“一种”和/或“该”等词并非特指单数，也可包括复数。一般说来，术语“包括”与“包含”仅提示包括已明确标识的步骤和元素，而这些步骤和元素不构成一个排它性的罗列，方法或者设备也可能包含其它的步骤或元素。

[20]本说明书中使用了流程图用来说明根据本说明书的实施例的系统所执行的操作。应

当理解的是，前面或后面操作不一定按照顺序来精确地执行。相反，可以按照倒序或同时处理各个步骤。同时，也可以将其他操作添加到这些过程中，或从这些过程移除某一步或数步操作。

[21]图 1 是根据本说明书的一个或多个实施例所示的知识图谱数据融合系统的应用场景示意图。

[22]知识图谱是指由一系列实体实例（即实体对应的数据实例）以及实体实例之间关系构成的知识库。其中，实体是对客观个体的广泛抽象，其可以指物理世界中的有形物体，如人、汽车、商户等，也可以指无形的对象，如话语、歌曲、电影、资金、程序代码等等。数据实例可以是实体的抽象概念下对应的实际存在的例子，如人可以具体是张三、李四、李明等，歌曲可以具体是《青花瓷》、《夜莺》、《天鹅湖》，商户具体可以是商户 A、商户 B、商户 C 等。实体实例之间可以具有关系，例如商户 A 与商户 B 有业务往来，商户 C 是商户 A 的子商户、张三是商户 A 的管理者等。在一些实施例中，实体实例之间的关系亦可看作是对应实体间的关系，例如，人与商户之间可以具有管理关系或雇佣关系等。在一些实施例中，知识图谱中的实体实例可以用节点表示，实体实例间关系可以用连接节点的边表示。

[23]知识图谱可以对应本体定义数据，或称为知识图谱的 schema。知识图谱的本体定义数据是指对知识图谱包括的实体、实体间关系进行定义的数据，可以表征知识图谱的本体的数据实例的语义信息。知识图谱的本体定义数据可以指导数据实例的收集，以及基于数据实例进行构图，得到知识图谱（也可以称为实例图）。因此，在一些实施例中，知识图谱的本体定义数据可以包括用于定义实体的实体字段。实体字段可以理解为实体名称或实体表征，如实体字段可以是“公司主体”、“用户”等，实体字段的取值则可以是前述的实体实例。实体字段可以对应多个属性字段，属性字段可以是对实体描述信息的抽象，如属性字段可以是“地址”、“年龄”、“注册资本”等，属性字段的取值则可以是其对应实体实例的具体描述，如“建设路 11 号”、“28 岁”、“500 万”等。在一些实施例中，知识图谱的本体定义数据可以包括用于定义实体间关系的关系描述，关系描述可以是实体间关系类型的抽象，如“雇佣关系”、“子母公司关系”、“父子关系”等。在一些实施例中，关系描述可以进一步包括关系属性，关系属性用于对关系描述做进一步说明，如“雇佣关系”可以具体是“临时雇佣”或者“正式雇佣”，“子母公司关系”可以进一步包括“全资控股关系”、“部分控股关系”等。通过关系描述可以在构建知识图谱时，确定两个实体实例之间是否具有边。在一些实施例中，还可以确定图谱算子。图谱算子用于基于实体定义或关系描述，从大量数据实例中找出实体实例及其确定实体实例之间的关系。图谱算子亦可理解为图计算算法或方法，用于进行图谱构建的数据处理操作或运算。可以用数据处理/运算单元、程序代码、机器学习模型等各种方式实现。在一些实施例中，可以对算子输入数据，算子可以进行相应的数据处理/运算，完成数据的转化，并输出转化后的数据。在一些实施例中，图谱算子可以看作是建立在知识图谱的本体定义数据（包括实体定义与关系描述）上的算法或方法，亦可看作是本体定义数据的一部分。

[24]本说明书提出的知识图谱数据融合系统可以应用在多平台或多业务领域数据处理的相关场景中，例如，可以应用在基于安全、保险、支付、财富等多个业务领域的数据进行业务任务（如确定某个自然人的资金风险）计算的场景。

[25]对于不同平台、不同业务领域，分别存储有各自的数据，例如各平台或业务领域可以以知识图谱或者数据表的形式记录各自的业务数据。通过不同平台、不同业务领域知识数据的融合和连通，可以提升业务效果、业务效率和计算效能。多平台、多业务领域的知识数据融合和连通可以通过构建多平台、多业务知识数据连通的知识图谱来实现。

[26]在一些实施例中，可以通过从各平台或各业务领域中获取数据表（即将数据实例通过二维表的形式记录，数据表中可以包括字段以及字段取值，即对应字段的数据实例，等），进一步基于获取的数据表创建（如构造图谱算子进行图谱计算）融合知识图谱。这种实施例涉及的构建融合知识图谱的方法基于不同平台或业务领域的知识数据实例来重新创建融合知识图谱，不能对不同各平台或不同业务领域已有的知识图谱加以利用，使得每一次数据融合时在构图过程中数据融合的实现成本高，数据维护的成本也高。另外，由于需要重新构图，开发周期较长，从各平台或各业务领域获取的知识数据实例，很可能需要存储在相应磁盘上以备使用，即各平台或各业务领域的知识数据会落入其它业务方的磁盘，无法保证数据安全。

[27]鉴于上述情况，本说明一些实施例提供了一种更加高效的知识图谱数据融合方法和系统，可以基于各平台或各业务领域已有的各个知识图谱的本体定义数据（如实体字段等实体定义数据、实体间关系描述等实体间关系定义数据）创建融合知识图谱的本体定义数据（如目标实体字段等实体定义数据、目标关系描述等实体间关系定义数据、对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子），再获取相关的各平台或各业务领域知识数据实例，根据融合知识图谱的本体定义数据对获取的知识数据实例进行处理得到融合知识图谱。通过本说明书一些实施例所述的知识图谱数据融合方法和系统，可以令融合知识图谱的构建实现自动化、标准化，构建过程更加高效，降低了知识数据融合、知识数据维护的成本。进一步的，本说明书一些实施例所述的知识图谱数据融合方法和系统可以在可信环境中执行，令各平台或各业务领域的知识数据（如知识数据实例）不落入其它业务方的磁盘，保护了知识数据隐私，保证了数据安全。

[28]在一些实施例中，本说明一些实施例提供的知识图谱数据融合方法和系统可以基于服务方、用户和业务方实现。用户可以是任何的个体或单位，例如个人、企业等。业务方可以是任何的个体或单位，业务方有与之对应的一个或多个的平台或业务领域，拥有各自的业务数据，在一些实施例中，业务方可以以知识图谱或者数据表的形式记录其业务数据。服务方可以是指用于实现知识图谱数据融合方法和系统的平台或系统，也可以是提供实现知识图谱数据融合方法和系统的平台或系统的任何个体或单位。在一些应用场景中，服务方可以基于一个或多个业务方（作为知识图谱提供方）的知识图谱向用户提供知识图谱数据融合服务。具体的，服务方可以获取来自一个或多个业务方的知识图谱的本体定义数据，并呈现给用户，用户可以在两个或更多个知识图谱的本体定义数据中确定其在融合服务中需要的实体字段以及关系描述，并可以将其作为目标实体字段以及目标关系描述指定（如告知或发送）给服务方。关于本体定义数据的具体内容可以参见图3及其相关描述。在一些实施例中，两个或更多个业务方的其中之一可以作为用户从服务方处请求并获得与其他业务方知识图谱数据相关的融合知识图谱数据。

[29]在一些实施例中，服务方可以获取目标实体字段以及目标关系描述，例如用户指定的目标实体字段以及目标关系描述，以及服务方还可以从两个或更多个知识图谱中获取

所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱。在一些实施例中，一个或多个用于对各目标实体字段以及各目标关系描述进行融合处理的图谱算子可以由服务方生成，也可以由用户生成并发送给服务方。服务方还可以通过目标任务算法处理融合知识图谱，得到目标任务结果并输出给用户。目标任务算法可以由服务方确定，或者还可以由用户向服务方指定。

[30]在一些实施例中，根据用户的权限，用户还可以从服务方处获取融合知识图谱的数据，例如，用户从相应的业务方处获得了数据使用权限，服务方可以验证用户的权限，如验证通过，则可将融合知识图谱发送给用户。

[31]如图 1 所示，知识图谱数据融合系统的应用场景 100 可以包括服务器 110-1、110-2、110-3 等多个服务器、处理设备 120 和网络 130。

[32]服务器 110-1、110-2、110-3 等多个服务器可以分别对应多个平台或业务领域。服务器 110-1、110-2、110-3、... 可以用于管理资源以及处理来自本系统至少一个组件或外部数据源（例如，云数据中心）的数据和/或信息。在一些实施例中，服务器 110-1、110-2、110-3、... 中的每一个可以是单一服务器或服务器组。该服务器组可以是集中式或分布式的（例如，服务器 110-1 可以是分布式系统），可以是专用的也可以由其他设备或系统同时提供服务。在一些实施例中，服务器 110-1、110-2、110-3、... 可以是区域的或者远程的。在一些实施例中，服务器 110-1、110-2、110-3、... 可以在云平台上实施，或者以虚拟方式提供。仅作为示例，所述云平台可以包括私有云、公共云、混合云、社区云、分布云、内部云、多层云等或其任意组合。

[33]服务器 110-1、110-2、110-3、... 中的任一个或更多个可以包括处理器 112。处理器 112 可以处理从其他设备或系统组成部分中获得的数据和/或信息。处理器可以基于这些数据、信息和/或处理结果执行程序指令，以执行一个或多个本申请中描述的功能。在一些实施例中，处理器 112 可以包含一个或多个子处理设备（例如，单核处理设备或多核多芯处理设备）。仅作为示例，处理器 112 可以包括中央处理器（CPU）、专用集成电路（ASIC）、专用指令处理器（ASIP）、图形处理器（GPU）、物理处理器（PPU）、数字信号处理器（DSP）、现场可编程门阵列（FPGA）、可编程逻辑电路（PLD）、控制器、微控制器单元、精简指令集电脑（RISC）、微处理器等或更多个任意组合。

[34]在一些实施例中，服务器 110-1、110-2、110-3、... 中的任一个或更多个可以存储对应平台或业务领域的的数据，例如数据实例、知识图谱的个体定义数据以及知识图谱等。在一些实施例中，服务器 110-1、110-2、110-3、... 中的任一个或更多个可以获取一个或更多个其它平台或业务领域的知识图谱的个体定义数据，还可以获取融合知识图谱的个体定义数据。在一些实施例中，服务器 110-1、110-2、110-3、... 可以对应不同的业务方。

[35]处理设备 120 可以处理从其他设备或系统组成部分中获得的数据和/或信息。处理设备 120 可以基于这些数据、信息和/或处理结果执行程序指令，以执行一个或多个本申请中描述的功能。在一些实施例中，处理设备 120 可以包含一个或多个子处理设备（例如，单核处理设备或多核多芯处理设备）。仅作为示例，处理设备 120 可以包括中央处理器（CPU）、专用集成电路（ASIC）、专用指令处理器（ASIP）、图形处理器（GPU）、物理处理器（PPU）、数字信号处理器（DSP）、现场可编程门阵列（FPGA）、可编程逻辑电路（PLD）、控制器、微控制器单元、精简指令集电脑（RISC）、微处理器等或更多个

任意组合。在一些实施例中，处理设备 120 可以属于所述服务方。

[36]网络 130 可以连接系统的各组成部分和/或连接系统与外部部分。网络 130 使得系统各组成部分之间以及与系统与外部部分之间可以进行通讯，促进数据和/或信息的交换。在一些实施例中，网络 130 可以是有线网络或无线网络中的任意一种或多种。例如，网络 130 可以包括电缆网络、光纤网络、电信网络、互联网、局域网（LAN）、广域网（WAN）、无线局域网（WLAN）、城域网（MAN）、公共交换电话网络（PSTN）、蓝牙网络、紫蜂网络（ZigBee）、近场通信（NFC）、设备内总线、设备内线路、线缆连接等或其任意组合。在一些实施例中，系统各部分之间的网络连接可以采用上述一种方式，也可以采取多种方式。在一些实施例中，网络 130 可以是点对点的、融合的、中心式的等各种拓扑结构或者多种拓扑结构的组合。在一些实施例中，网络 130 可以包括一个或更多个网络接入点。例如，网络 130 可以包括有线或无线网络接入点，例如基站和/或网络交换点 130-1、130-2、...，通过这些网络接入点，系统 100 的一个或多个组件可连接到网络 130 以交换数据和/或信息。

[37]在一些实施例中，处理设备 120 可以通过网络 130 从服务器 110-1、110-2、110-3、... 中的两个或更多个获取两个或更多个知识图谱的本体定义数据（如实体字段等实体定义数据、实体间关系描述等实体间关系定义数据）创建融合知识图谱的本体定义数据（如目标实体字段等实体定义数据、目标关系描述等实体间关系定义数据、对各目标实体字段以及各目标关系描述进行融合处理的图谱算子），再通过网络 130 从服务器 110-1、110-2、110-3、... 中获取相关的各平台或各业务领域数据实例，根据融合知识图谱的本体定义数据对获取的数据实例进行处理得到融合知识图谱。在一些实施例中，处理设备 120 可以是用于实现知识图谱数据融合的专用设备，用于接收来自用户（图中未示出）或其他平台或业务领域（如服务器 110-1、110-2、110-3、... 中的任意一个或更多个）的数据融合请求，并返回融合数据。在一些实施例中，用户或服务器 110-1、110-2、110-3、... 中的任意一个或更多个还可以将目标任务和/或目标任务算法通过网络 130 发送至处理设备 120，处理设备 120 可以通过目标任务和/或目标任务算法处理融合知识图谱，得到目标任务结果并输出，用户或服务器 110-1、110-2、110-3、... 中的任意一个或更多个可以通过网络 130 接受处理设备 120 输出的目标任务结果。在一些实施例中，处理设备 120 可以部署在服务器 110-1、110-2、110-3、... 中的某一个上，或者服务器 110-1、110-2、110-3、... 中的某一个可以作为处理设备 120 以实现处理设备 120 的功能。换句话说，在一些应用场景中，业务方还可以作为服务方，提供知识图谱数据融合服务。

[38]图 2 是根据本说明书一些实施例所示的一种知识图谱数据融合系统的框图。

[39]在一些实施例中，知识图谱数据融合系统 200 可以实现于服务器 110-1、110-2、110-3、... 中的某一个上或处理设备 120 上。其可以包括目标数据获取模块 210、图谱算子确定模块 220 和融合图谱生成模块 230。在一些实施例中，知识图谱数据融合系统 200 还可以包括展示模块 240。在一些实施例中，知识图谱数据融合系统 200 还可以包括图谱处理模块 250。

[40]在一些实施例中，目标数据获取模块 210 可以用于获取目标实体字段以及目标关系描述；所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据；其中，知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系

的关系描述。

[41]在一些实施例中，图谱算子确定模块 220 可以用于确定一个或多个用于对各目标实体字段以及各目标关系描述进行融合处理的图谱算子。在一些实施例中，所述实体字段对应有一个或多个属性字段。在一些实施例中，所述图谱算子用于实现以下操作中的一种或多种：将目标实体字段对应的属性字段的实例值进行表达标准化处理；将两个或更多的目标实体字段进行融合，以得到融合实体字段；融合实体字段对应的属性字段来自所述两个或更多的目标实体字段中的至少一个对应的属性字段；融合实体字段相关的关系描述包括所述两个或更多的目标实体字段中的每一个相关的目标关系描述；基于两个目标实体字段中的至少一个对应的属性字段，建立相应两个目标实体间的关系描述；以及，调用自然语言处理模型确定数据实例中相似实例，以便将数据实例中的相似实例进行融合。

[42]在一些实施例中，融合图谱生成模块 230 可以用于从两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱。在一些实施例中，融合图谱生成模块 230 还可以用于确定图谱算子涉及的目标实体字段以及目标关系描述，作为最小子图的实体字段和关系描述；从各知识图谱中获取最小子图的实体字段和关系描述对应的数据实例；通过图谱算子处理最小子图的实体字段和关系描述对应的数据实例，得到最小子图；从各知识图谱中获取最小子图的实体字段和关系描述以外的目标实体字段以及目标关系描述对应的数据实例，得到融合知识图谱除最小子图以外的子图。

[43]在一些实施例中，展示模块 240 可以用于基于所述目标实体字段、所述目标关系描述以及所述图谱算子获取融合知识图谱的本体定义数据，以及通过知识图谱视图的形式表达所述融合知识图谱的本体定义数据。

[44]在一些实施例中，图谱处理模块 250 可以用于通过目标任务算法处理所述融合知识图谱，得到目标任务结果并输出；所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

[45]在一些实施例中，融合图谱生成模块 230 可以部署在可信执行环境中。

[46]在一些实施例中，图谱处理模块 250 可以部署在可信执行环境中。

[47]应当理解，所示的系统及其模块可以利用各种方式来实现。例如，在一些实施例中，系统及其模块可以通过硬件、软件或者软件和硬件的结合来实现。其中，硬件部分可以利用专用逻辑来实现；软件部分则可以存储在存储器中，由适当的指令执行系统，例如微处理器或者专用设计硬件来执行。本领域技术人员可以理解上述的方法和系统可以使用计算机可执行指令和/或包含在处理器控制代码中来实现，例如在诸如磁盘、CD 或 DVD-ROM 的载体介质、诸如只读存储器（固件）的可编程的存储器或者诸如光学或电子信号载体的数据载体上提供了这样的代码。本说明书的系统及其模块不仅可以有诸如超大规模集成电路或门阵列、诸如逻辑芯片、晶体管等的半导体、或者诸如现场可编程门阵列、可编程逻辑设备等的可编程硬件设备的硬件电路实现，也可以用例如由各种类型的处理器所执行的软件实现，还可以由上述硬件电路和软件的结合（例如，固件）来实现。

[48]需要注意的是，以上对于系统及其模块的描述，仅为描述方便，并不能把本说明书

限制在所举实施例范围之内。可以理解，对于本领域的技术人员来说，在了解该系统的原理后，可能在不背离这一原理的情况下，对各个模块进行任意组合，或者构成子系统与其他模块连接。

[49]图 3 是根据本说明书一些实施例所示的一种知识图谱数据融合方法的示例性流程图。

[50]在一些实施例中，方法 300 可以由处理设备 120 执行。在一些实施例中，方法 300 可以由部署于处理设备 120 上的知识图谱数据融合系统 200 实现。

[51]如图 3 所示，该方法 300 可以包括以下步骤。

[52]步骤 310，获取目标实体字段以及目标关系描述，所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据。

[53]在一些实施例中，该步骤 310 可以由目标数据获取模块 210 执行。

[54]在一些实施例中，两个或更多个的知识图谱的本体定义数据可以来自于两个或更多个的平台或业务领域，两个或更多个的平台或业务领域可以对应属于一个或多个知识图谱提供方，例如业务方。在一些实施例中，不同平台或业务领域的知识图谱的数据表达标准可以不同。例如，属性字段的格式可以不同，或者同一实体在不同平台或业务领域的知识图谱 schema 中被定义的实体字段不同，如实体为公司，在业务领域 A 中的 schema 定义为实体字段“CRO.company”，在业务领域 B 中的 schema 定义为实体字段“CompanyV2”。

[55]知识图谱的本体定义数据可以可视化呈现，关于知识图谱的本体定义数据的可视化示意图及更多内容可以参见图 4 及其相关描述。

[56]在一些实施例中，目标数据获取模块 210 可以根据业务目标等实际需求从两个或更多个平台/业务领域的知识图谱的本体定义数据中筛选出所需的实体字段以及关系描述，被选中的实体字段以及关系描述被称为目标实体字段以及目标关系描述。例如，业务目标是判断商户的资金风险，则可以从保险业务领域的知识图谱本体定义数据中筛选出商户、商品、投保人、管理者等与商户有关的实体字段作为目标实体字段和属于、管理、投保等相关的关系描述作为目标关系描述，以及可以从支付业务领域的知识图谱本体定义数据中筛选出商家、商品、收款人、管理者等与商户有关的实体字段作为目标实体字段和属于、管理、支付等相关的关系描述作为目标关系描述。在一些实施例中，从同一个知识图谱的本体定义数据选出的关系描述应当是与同时选出的实体字段相关的。换句话说，从知识图谱本体定义数据中筛选出的关系描述涉及的实体字段都在被选中的目标实体字段中。相反，从知识图谱本体定义数据中筛选出的实体字段涉及的关系描述则可以不在被选中的目标关系描述中。

[57]在一些实施例中，可以由用户从两个或更多个平台/业务领域的知识图谱的本体定义数据中筛选出目标实体字段以及目标关系描述。

[58]步骤 320，确定一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子。

[59]在一些实施例中，该步骤 320 可以由图谱算子确定模块 220 执行。

[60]为了构建融合知识图谱，可以确定用于对各目标实体字段以及各目标关系描述进行融合处理的图谱算子。关于图谱算子的一般性描述可以参见以上的相关说明。用于融合处理的图谱算子是指用于实现各目标实体字段以及各目标关系描述所对应数据的融合

和/或连通处理的各种图谱算子。例如,可以包括将相似的目标实体融合为一个实体、给未关联的两个目标实体间加上关系、对属性信息进行表达标准化处理等各种算子。关于图谱算子的更多内容可以参见步骤 330 及其相关描述。

[61]在一些实施例中,处理设备 120 可以自行生成或者也可以由用户生成用于对各目标实体字段以及各目标关系描述进行融合处理的图谱算子,并由用户提供给处理设备 120。

[62]可以理解,对于不同平台/业务领域的知识图谱,其包括的本体定义数据可以不同,即目标实体字段以及目标关系描述可以不同,且不同平台/业务领域的知识图谱的本体定义数据之间并未连通,如各目标实体字段之间并未关联。通过确定一个或更多个用于对各目标实体字段以及各目标关系描述进行融合处理的图谱算子,可以将不同平台/业务领域的知识图谱的本体定义数据进行融合和关联,得到用于构建融合知识图谱的本体定义数据,进而可以基于融合知识图谱的本体定义数据实现不同平台/业务领域知识图谱对应的数据实例的融合和/或连通。

[63]知识图谱的本体定义数据为数据实例的概括或抽象,将其公开融合给各个平台/各个业务领域,并不会导致敏感数据实例的公开。

[64]在一些实施例中,融合知识图谱的本体定义数据可以通过知识图谱视图的形式表达。知识图谱视图可以在展示界面(例如终端界面)中进行图形显示,例如用节点表示目标实体字段或融合后的实体字段,用连接两个节点的边表示实体间的关系描述,包括基于图谱算子新建的关系描述,在一些实施例中,处理设备 120 可以向用户发送融合知识图谱的本体定义数据,以便向用户直观呈现即将生成的知识图谱的本体框架,进而方便用户对融合知识图谱的本体定义数据进行调整或完善,提高构图效率。关于知识图谱视图的更多内容可以参见图 4 及其相关描述,此处不再赘述。

[65]步骤 330,从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例,并通过所述图谱算子处理所述数据实例以生成融合知识图谱。

[66]在一些实施例中,该步骤 330 可以由融合图谱生成模块 230 执行。

[67]在一些实施例中,确定得到融合知识图谱的本体定义数据后,即可以根据融合知识图谱的本体定义数据,如目标实体字段、目标关系描述以及目标实体字段对应的属性字段从对应的平台或业务领域的知识图谱中获取相应的数据实例,数据实例可以包括目标实体字段对应的实体实例、属性值以及这些实体实例之间的关系描述。

[68]在一些实施例中,一个或多个图谱算子实现的数据处理操作/运算可以包括将目标实体字段对应的属性字段的实例值进行表达标准化处理。表达标准化处理可以是将属性字段的实例值的数据格式(如属性字段的实例值为数值或字符或二进制数)、数据表达约束条件(如时间类型的属性字段约束条件为实例值为年月日的取值或 24 小时时间类型的取值,金额类型的属性字段约束条件为实例值为以美元为单元的取值或以人民币为单位的取值)、数据表达类型(如属性字段的实例值为整型数据或浮点型数据)等进行统一的标准化处理,使来自不同平台或业务领域的实体字段的属性值具有统一的表达形式或度量方式。

[69]在一些实施例中,一个或多个图谱算子实现的数据处理操作/运算可以包括将两个或更多个的目标实体字段进行融合,以得到融合实体字段。可以理解,可以基于不同知识图谱的本体定义数据的融合,如通过将两个或更多个的目标实体字段进行融合,来实现

不同平台/业务领域的知识数据的融合和连通。在一些实施例中，可以将语义相似或相同的目标实体字段进行融合。例如融合知识图谱中的本体定义数据中包括来自保险业务领域的目标实体字段“CRO.company”和来自支付业务领域的目标实体字段“CompanyV2”，可以对“CRO.Company”和“CompanyV2”进行融合，得到融合实体字段，融合实体字段可以用被融合的两个或更多个目标实体字段中的任一个来表示如“CRO.Company”或“CompanyV2”，也可以用其它能够表达被融合的两个或更多个目标实体字段的语义的其它实体字段来表示。在一些实施例中，两个或更多个目标实体字段融合得到融合实体字段后，被融合的两个或更多个目标实体字段对应的属性字段、相关的关系描述也会被调整以便与融合实体字段相适应。具体的，融合实体字段对应的属性字段可以是被融合的两个或更多个的目标实体字段对应的属性字段的并集，或者是所述并集中的一部分，例如，融合实体字段对应的属性字段可以是被融合的某一个目标实体字段对应的全部或部分属性字段等等。融合实体字段相关的关系描述可以包括被融合的两个或更多个的目标实体字段中的每一个相关的目标关系描述。

[70]在一些实施例中，可以计算各个目标实体字段之间的相似度，将相似度满足条件（如相似度大于阈值或相似度排名为 TopN）的两个或更多个的目标实体字段进行融合以得到融合实体字段。

[71]在一些实施例中，可以通过 tf-idf 算法、计算文本间的向量距离（距离可以包括但不限于余弦距离、欧式距离、曼哈顿距离、马氏距离或闵可夫斯基距离等）等文本相似度算法计算目标实体字段之间的相似度。

[72]在一些实施例中，可以通过语义相似度预测模型确定两个目标实体字段的相似度，例如可以基于 BERT、Transformer、ESIM 等模型计算目标实体字段之间的相似度。在一些实施例中，还可以基于目标实体字段对应的属性字段来确定两个或更多个的目标实体字段是否相似或相同。以 BERT 模型为例，可以将两个或更多个目标实体字段对应的文本（可以包括目标实体字段的字段名以及对应的属性字段名）输入 BERT 模型，BERT 模型可以确定两个或更多个目标实体字段的文本向量，并计算文本向量之间的语义相似度，BERT 模型可以输出文本向量之间的相似度评分，即可以将得到的相似度评分作为目标实体字段之间的相似度。

[73]以融合图谱算子：fusion(CRO.Company,CompanyV2)为例，其定义将来自不同平台的知识图谱 schema 中的目标实体字段 CRO.Company 以及目标实体字段 CompanyV2 进行融合，该图谱算子可以对应一段程序代码，在基于数据实例生成融合知识图谱时被调用，将“CRO.Company”对应的实体实例和“CompanyV2”对应的实体实例处理为同一个实体字段，即融合实体字段，下的实例。

[74]在一些实施例中，一个或多个图谱算子实现的数据处理操作/运算可以包括基于两个目标实体字段中的至少一个对应的属性字段，建立相应两个目标实体间的关系描述。如前所述，实体字段对应的属性字段可以表示实体字段的进一步描述信息的定义，如名称、地址、类型等，在一些实施例中，由目标实体字段对应的属性字段可以确定未关联的两个目标实体之间是否存在新的关联关系，进而可以建立两个目标实体间的关系描述。例如，来自保险业务领域的目标实体字段“CRO.Company”对应的属性字段包括“address”，目标实体字段“City”来自支付业务领域，可以根据“CRO.Company”对应的属性字段

“address”，建立“CRO.Company”和“City”间的关系描述，如建立关系描述为所在城市。又例如，来自制造业务领域的目标实体字段“商品”对应有属性字段“商品类型”，来自销售业务领域的目标实体字段“商户”也对应有属性字段“主营范围”，则可以基于两者的属性字段建立“商品”和“商户”之间的关系描述，如建立关系描述为销售关系。

[75]以链指图谱算子：link(CRO.Company,inCity,City,address)为例，其可以基于目标实体字段“CRO.Company”、“CRO.Company”的属性字段“address”、以及目标实体字段“city”定义“CRO.Company”和“City”间的关系描述。调用该算子处理目标实体字段“CRO.Company”和“City”以及“CRO.company”和“City”对应的数据实例时，可以基于“CRO.Company”实体实例属性字段“address”值，建立“CRO.Company”实体实例和“City”实体实例间的关系描述为“inCity”。

[76]在一些实施例中，一个或多个图谱算子实现的数据处理操作/运算还可以包括确定数据实例中的相似实例，以便将数据实例中的相似实例进行融合。例如融合实体字段对应的数据实例中包括相似的2个数据实例“酒店D”和“快捷酒店D”，则可以通过图谱算子将“酒店D”和“快捷酒店D”进行融合，得到融合后的数据实例，如融合得到“酒店D”。在一些实施例中，可以在图谱算子中加入用于调用自然语言处理模型进行数据处理的接口调用代码，从而实现调用自然语言处理模型实现前述的数据处理。

[77]在一些实施例中，调用自然语言处理模型确定数据实例中的相似实例可以通过调用自然语言处理模型确定数据实例的实体字段取值和/或其属性字段值的相似度，并确定相似度满足条件（如相似度大于阈值或相似度排名为TopN）的两个或更多个的数据实例作为相似实例。在一些实施例中，自然语言模型可以是用于自然语言处理的神经网络模型例如BERT、Transformer、ESIM等模型，可以采用与确定目标实体字段之间相似度类似的方法来实现通过神经网络模型处理数据实例的实体字段取值和/或其属性字段值，得到数据实例之间的相似度，此处不再赘述。

[78]在一些实施例中，通过确定好的图谱算子对图谱算子涉及的目标实体字段以及目标关系描述对应的数据实例进行处理，即可以生成融合知识图谱。

[79]在一些实施例中，生成融合知识图谱后，可以根据业务目标任务（如判断商户的资金风险）处理融合知识图谱，得到目标任务结果（如商户资金风险类型为中高风险）并输出给业务方或其它用户，以实现基于多平台/多业务领域连通的知识数据来更加高效、准确地进行业务任务计算。

[80]在一些实施例中，方法300还可以包括步骤340：通过目标任务算法处理融合知识图谱以得到目标任务结果并输出。在一些实施例中，步骤340可以由图谱处理模块250执行。

[81]目标任务算法可以指用于进行目标任务计算的各种算法，例如可以包括图谱规则推理算法、基于图谱的机器学习模型预测算法等。

[82]图谱规则推理算法是指基于知识图谱的实体实例、实体实例间关系等知识数据进行规则推理来得到目标任务的结果的算法，例如基于融合知识图谱查询/推理两个或更多个实例的关系，如李四的亲属有哪些、某管理者管理的商户有哪些等。

[83]基于图谱的机器学习模型预测算法是指通过机器学习模型处理知识图谱来实现目标任务的结果预测的算法，例如基于图卷积网络处理融合知识图谱，得到融合知识图谱的

表达如实体对应的向量表示，再基于表达对融合知识图谱中的实体进行分类，即得到融合知识图谱的某些实体属于哪一类的预测结果。

[84]在一些实施例中，目标任务算法可以由处理设备 120（即服务方）自行确定，也可以由用户指定。

[85]在一些实施例中，本说明书的一些实施例所示的知识图谱数据融合方法中的至少部分步骤在可信环境中执行，例如从各知识图谱中获取目标实体字段以及目标关系描述对应的数据实例，并通过图谱算子处理数据实例以生成融合知识图谱，又例如根据业务目标任务处理融合知识图谱，得到目标任务结果。

[86]在一些实施例中，可信环境可以是可信执行环境 TEE(Trusted Execution Environment) 或支持全内存计算的设备内存等能够将其中的数据与外界隔离的执行环境。例如，外界不能访问可信环境中的数据，也不能控制在其中执行的代码。全内存计算是指数据事先存储于内存，计算过程中直接从内存读写数据，计算产生的中间结果也不落磁盘。

[87]在一些实施例中，通过图谱算子处理数据实例以生成融合知识图谱，又例如根据业务目标任务处理融合知识图谱，得到目标任务结果都可以基于全内存计算进行。

[88]在一些实施例中，在可信执行环境中执行的各个方法步骤生成的中间结果，在计算完成后都可以被销毁，例如从各知识图谱中获取的目标实体字段以及目标关系描述对应的数据实例、通过图谱算子处理数据实例生成的融合知识图谱、根据业务目标任务处理融合知识图谱的中间结果等。

[89]通过在可信执行环境中执行知识图谱数据融合方法的至少部分步骤或将知识图谱数据融合系统 200 中的一个或多个模块部署在可信执行环境中，可以实现各平台/各业务领域的的数据实例不落入其它业务方的磁盘，在实现数据高效融合的同时保证各方数据的安全和隐私。

[90]在一些实施例中，处理设备 120 可以根据用户权限，向用户输出融合知识图谱或者目标任务结果，以此从服务方处获取知识图谱融合服务。

[91]图 4 是根据本说明书一些实施例所示的一种融合知识图谱的本体定义数据的可视化 400 示意图。

[92]图 4 所示的融合知识图谱的本体定义数据可以知识图谱视图（如 KGView）的形式在展示界面（如系统、平台、应用程序等界面上）上展示。在一些实施例中，融合知识图谱的本体定义数据的可视化过程可由展示模块 240 实现。

[93]如图 4 中所示，展示了两个业务领域 A 和 B 的知识图谱本体定义数据对应的 2 个知识图谱视图 a 和 b，以及由两个业务领域的本体定义数据得到的融合知识图谱的本体定义数据对应的知识图谱视图 c。在知识图谱视图中，实体字段用节点表示（在图 4 中，圆圈表示节点），实体间关系描述用连接两个节点的边表示（在图 4 中，圆圈与圆圈之间的连线称为边）。

[94]如图 4 所示，可以从业务领域 A 的知识图谱本体定义数据中选取目标实体字段“机械”和“商户”以及目标关系描述“所售产品”，从业务领域 B 的知识图谱本体定义数据中选取目标实体字段“机具”和“小程序”，并确定用于将“机械”和“机具”融合的图谱算子和用于建立“商户”和“小程序”关系描述为“收款途径”的图谱算子。基于选择的目标实体字段、目标关系描述、确定的图谱算子，得到融合知识图谱的本体定义数据对应的知识图

谱视图 c, 其中“机具”为“机械”和“机具”融合得到的融合实体字段, “收款途径”为在“商户”和“小程序”之间建立的边表示的关系描述。

[95]在通过图谱算子处理所述目标实体字段以及目标关系描述对应的数据实例, 生成融合知识图谱的过程中, 为了进一步提高融合知识图谱的生成效率, 降低运算成本或运算开销, 本说明书另一些实施例提供了一种生成融合知识图谱的方法。

[96]图 5 是根据本说明书另一些实施例所示的一种生成融合知识图谱的示例性流程图。

[97]在一些实施例中, 方法 500 可以由处理设备 120 执行。在一些实施例中, 方法 500 可以由部署于处理设备 120 上的融合图谱生成模块 230 实现。

[98]如图 5 所示, 该方法 500 可以包括以下步骤。

[99]步骤 510, 确定图谱算子涉及的目标实体字段以及目标关系描述, 作为最小子图的实体字段和关系描述。

[100]如前所述, 图谱算子用于对目标实体字段以及目标关系描述进行融合处理, 即图谱算子中包括了需要进行融合处理的目标实体字段以及目标关系描述。在一些实施例中, 在融合知识图谱的本体定义数据中, 可以只需要对部分目标实体字段以及目标关系描述进行融合处理。为了节约计算资源, 即可以在融合知识图谱的本体定义数据中确定图谱算子涉及的目标实体字段以及目标关系描述, 并将该部分目标实体字段以及目标关系描述, 作为最小子图的实体字段和关系描述。其中, 最小子图是指基于图谱算子涉及的目标实体字段以及目标关系描述对应的数据实例构建的知识图谱子图。在一些实施例中, 可以将融合知识图谱的本体定义数据中全部图谱算子涉及的目标实体字段以及目标关系描述作为最小子图的实体字段和关系描述, 换句话说, 一个融合知识图谱对应一个最小子图。在一些实施例中, 可以将融合知识图谱的本体定义数据中不同图谱算子分别涉及的目标实体字段以及目标关系描述作为不同最小子图的实体字段和关系描述, 换句话说, 一个融合知识图谱对应多个最小子图。

[101]步骤 520, 从各知识图谱中获取最小子图的实体字段和关系描述对应的数据实例。

[102]在一些实施例中, 确定一个或多个最小子图包括的实体字段和关系描述后, 便可以从各个知识图谱中获取最小子图的实体字段和关系描述对应的数据实例, 如图 5 中业务域 A、B 中白色子图, 将其用于后续的融合知识图谱生成处理。相较于获取融合知识图谱的所有目标实体字段和目标关系描述对应的数据实例来进行后续处理, 通过本实施例可以提高融合知识图谱的数据处理效率。

[103]步骤 530, 通过图谱算子处理最小子图的实体字段和关系描述对应的数据实例, 得到最小子图。

[104]在一些实施例中, 通过图谱算子处理最小子图的实体字段和关系描述对应的数据实例, 可以得到融合了部分目标实体字段和目标关系描述对应的数据实例的最小子图, 如图 5 中融合知识图谱中的白色子图。

[105]在一些实施例中, 可以通过多个图谱算子分别处理多个最小子图对应的实体字段和关系描述对应的数据实例, 得到多个最小子图。

[106]步骤 540, 从各知识图谱中获取最小子图的实体字段和关系描述以外的目标实体字段以及目标关系描述对应的数据实例, 得到融合知识图谱除最小子图以外的子图。

[107]在一些实施例中, 完成融合知识图谱的图谱算子所涉及的一个或多个最小子图的生

成后,便完成了融合知识图谱中需要进行融合处理的目标实体字段和目标关系描述对应的数据实例的融合处理,即实现了多平台/多业务领域的知识数据连通。

[108]得到了融合知识图谱的各个最小子图后,可以从各平台/各业务领域的各知识图谱中获取最小子图的实体字段和关系描述以外的目标实体字段以及目标关系描述对应的数据实例,如图5中业务域A、B中的灰色子图,得到融合知识图谱中除最小子图以外的子图,如图5中融合知识图谱中的灰色子图,将最小子图和除最小子图以外的子图加载在一起,即得到了包括完整知识数据的融合知识图谱。

[109]可以理解的,最小子图的实体字段和关系描述对应的数据实例是融合知识图谱中的部分且需要进行融合处理的数据实例,融合知识图谱中的其余部分实体字段和关系描述对应的数据实例及其之间的关系则可以直接从已有的各知识图谱中直接获取。通过本实施例,可以充分利用已有知识图谱的知识数据,显著降低生成融合知识图谱的计算成本。

[110]在一些实施例中,用户可以向服务方请求知识图谱融合服务,并从服务方处获取融合数据。在一些实施例中,用户还可以进行定制化要求,如指定目标实体字段、目标关系描述以及用于处理融合知识图谱的目标任务算法。图6是根据本说明书另一些实施例所示的一种知识图谱数据处理方法的示例性流程图。

[111]在一些实施例中,用户可以通过终端等设备实现方法600中的一个或多个步骤。

[112]如图6所示,该方法600可以包括以下步骤。

[113]步骤610,向服务方指定目标实体字段以及目标关系描述。

[114]在一些实施例中,用户可以从两个或更多个知识图谱的本体定义数据中筛选目标实体字段以及目标关系描述,并将目标实体字段以及目标关系描述指定给服务方。其中,两个或更多个的知识图谱的本体定义数据可以来自于两个或更多个的平台或业务领域,两个或更多个的平台或业务领域可以对应属于一个或多个知识图谱提供方例如业务方。关于知识图谱的本体定义数据、目标实体字段以及目标关系描述的更多内容可以参见步骤310及其相关描述。

[115]步骤620,从所述服务方处获取融合知识图谱和/或从所述服务方处获取目标任务结果。

[116]在一些实施例中,服务方可以通过方法300获取融合知识图谱以及目标任务结果,并将融合知识图谱和/或目标任务结果发送给用户。

[117]在一些实施例中,用户还可以从所述服务方获取通过知识图谱视图的形式表达的融合知识图谱的本体定义数据。关于知识图谱视图的形式表达的融合知识图谱的本体定义数据的更多内容可以参见图4及其相关说明。

[118]本说明书另一个方面提供一种知识图谱数据处理系统。

[119]在一些实施例中,知识图谱数据处理系统可以包括目标数据指定模块和结果获取模块。

[120]在一些实施例中,目标数据指定模块可以用于向服务方指定目标实体字段以及目标关系描述;所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据;其中,知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述。

[121]在一些实施例中,知识图谱数据处理系统还可以包括算子确定模块,可以用于生成

一个或多个用于对各目标实体字段以及各目标关系描述进行融合处理的图谱算子，并发送给所述服务方。

[122]在一些实施例中，知识图谱数据处理系统还可以包括算法确定模块，可以用于向所述服务方指定目标任务算法。

[123]在一些实施例中，结果获取模块可以用于从所述服务方处获取融合知识图谱和/或从所述服务方处获取目标任务结果；所述融合知识图谱通过图谱算子处理数据实例生成，所述数据实例基于所述目标实体字段以及所述目标关系描述从所述两个或更多个知识图谱中获取；所述目标任务结果通过目标任务算法处理所述融合知识图谱得到；所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

[124]在一些实施例中，结果获取模块还可以用于从所述服务方获取通过知识图谱视图的形式表达的融合知识图谱的本体定义数据；所述融合知识图谱的本体定义数据基于所述目标实体字段、所述目标关系描述以及所述图谱算子获取。

[125]应当理解，所示的系统及其模块可以利用各种方式来实现。例如，在一些实施例中，系统及其模块可以通过硬件、软件或者软件和硬件的结合来实现。其中，硬件部分可以利用专用逻辑来实现；软件部分则可以存储在存储器中，由适当的指令执行系统，例如微处理器或者专用设计硬件来执行。本领域技术人员可以理解上述的方法和系统可以使用计算机可执行指令和/或包含在处理器控制代码中来实现，例如在诸如磁盘、CD 或 DVD-ROM 的载体介质、诸如只读存储器（固件）的可编程的存储器或者诸如光学或电子信号载体的数据载体上提供了这样的代码。本说明书的系统及其模块不仅可以有诸如超大规模集成电路或门阵列、诸如逻辑芯片、晶体管等的半导体、或者诸如现场可编程门阵列、可编程逻辑设备等的可编程硬件设备的硬件电路实现，也可以用例如由各种类型的处理器所执行的软件实现，还可以由上述硬件电路和软件的结合（例如，固件）来实现。

[126]需要注意的是，以上对于系统及其模块的描述，仅为描述方便，并不能把本说明书限制在所举实施例范围之内。可以理解，对于本领域的技术人员来说，在了解该系统的原理后，可能在不背离这一原理的情况下，对各个模块进行任意组合，或者构成子系统与其他模块连接。

[127]本说明书实施例还提供一种知识图谱数据融合装置，包括至少一个存储介质和至少一个处理器，所述至少一个存储介质用于存储计算机指令；所述至少一个处理器用于执行所述计算机指令以实现所述的知识图谱数据融合方法。

[128]本说明书另一个方面提供一种知识图谱数据处理装置，包括至少一个存储介质和至少一个处理器，所述至少一个存储介质用于存储计算机指令；所述至少一个处理器用于执行所述计算机指令以实现所述的知识图谱数据处理方法。

[129]本说明书实施例可能带来的有益效果包括但不限于：（1）基于各平台或各业务领域已有的各个知识图谱的本体定义数据创建融合知识图谱的本体定义数据，再获取相关的各平台或各业务领域数据实例，根据融合知识图谱本体定义数据中用于对不同平台或业务领域的实体字段和关系描述进行融合处理的图谱算子对获取的数据实例进行处理，生成融合知识图谱，可以令融合知识图谱的构建实现自动化、标准化，构建过程更加高效，降低了数据融合、数据维护的成本；（2）知识图谱数据融合方法可以在可信环境中执行，

提高数据融合效率的同时有效保护了数据隐私；(3) 基于最小子图的融合知识图谱生成方法可以充分利用已有知识图谱的知识数据，进一步降低计算成本。需要说明的是，不同实施例可能产生的有益效果不同，在不同的实施例里，可能产生的有益效果可以是以上任意一种或几种的组合，也可以是其他任何可能获得的有益效果。

[130]上文已对基本概念做了描述，显然，对于本领域技术人员来说，上述详细披露仅作为示例，而并不构成对本说明书的限定。虽然此处并没有明确说明，本领域技术人员可能会对本说明书进行各种修改、改进和修正。该类修改、改进和修正在本说明书中被建议，所以该类修改、改进、修正仍属于本说明书示范实施例的精神和范围。

[131]同时，本说明书使用了特定词语来描述本说明书的实施例。如“一个实施例”、“一实施例”、和/或“一些实施例”意指与本说明书至少一个实施例相关的某一特征、结构或特点。因此，应强调并注意的是，本说明书中在不同位置两次或多次提及的“一实施例”或“一个实施例”或“一个替代性实施例”并不一定是指同一实施例。此外，本说明书的一个或多个实施例中的某些特征、结构或特点可以进行适当的组合。

[132]此外，本领域技术人员可以理解，本说明书的各方面可以通过若干具有可专利性的种类或情况进行说明和描述，包括任何新的和有用的工序、机器、产品或物质的组合，或对他们的任何新的和有用的改进。相应地，本说明书的各个方面可以完全由硬件执行、可以完全由软件（包括固件、常驻软件、微码等）执行、也可以由硬件和软件组合执行。以上硬件或软件均可被称为“数据块”、“模块”、“引擎”、“单元”、“组件”或“系统”。此外，本说明书的各方面可能表现为位于一个或多个计算机可读介质中的计算机产品，该产品包括计算机可读程序编码。

[133]计算机存储介质可能包含一个内含有计算机程序编码的传播数据信号，例如在基带上或作为载波的一部分。该传播信号可能有多种表现形式，包括电磁形式、光形式等，或合适的组合形式。计算机存储介质可以是除计算机可读存储介质之外的任何计算机可读介质，该介质可以通过连接至一个指令执行系统、装置或设备以实现通讯、传播或传输供使用的程序。位于计算机存储介质上的程序编码可以通过任何合适的介质进行传播，包括无线电、电缆、光纤电缆、RF、或类似介质，或任何上述介质的组合。

[134]本说明书各部分操作所需的计算机程序编码可以用任意一种或多种程序语言编写，包括面向对象编程语言如 Java、Scala、Smalltalk、Eiffel、JADE、Emerald、C++、C#、VB.NET、Python 等，常规程序化编程语言如 C 语言、Visual Basic、Fortran2003、Perl、COBOL2002、PHP、ABAP，动态编程语言如 Python、Ruby 和 Groovy，或其他编程语言等。该程序编码可以完全在用户计算机上运行、或作为独立的软件包在用户计算机上运行、或部分在用户计算机上运行部分在远程计算机运行、或完全在远程计算机或处理设备上运行。在后种情况下，远程计算机可以通过任何网络形式与用户计算机连接，比如局域网（LAN）或广域网（WAN），或连接至外部计算机（例如通过因特网），或在云计算环境中，或作为服务使用如软件即服务（SaaS）。

[135]此外，除非权利要求中明确说明，本说明书所述处理元素和序列的顺序、数字字母的使用、或其他名称的使用，并非用于限定本说明书流程和方法的顺序。尽管上述披露中通过各种示例讨论了一些目前认为有用的发明实施例，但应当理解的是，该类细节仅起到说明的目的，附加的权利要求并不仅限于披露的实施例，相反，权利要求旨在覆盖

所有符合本说明书实施例实质和范围的修正和等价组合。例如，虽然以上所描述的系统组件可以通过硬件设备实现，但是也可以只通过软件的解决方案得以实现，如在现有的处理设备或移动设备上安装所描述的系统。

[136]同理，应当注意的是，为了简化本说明书披露的表述，从而帮助对一个或多个发明实施例的理解，前文对本说明书实施例的描述中，有时会将多种特征归并至一个实施例、附图或对其的描述中。但是，这种披露方法并不意味着本说明书对象所需要的特征比权利要求中提及的特征多。实际上，实施例的特征要少于上述披露的单个实施例的全部特征。

[137]一些实施例中使用了描述成分、属性数量的数字，应当理解的是，此类用于实施例描述的数字，在一些示例中使用了修饰词“大约”、“近似”或“大体上”来修饰。除非另外说明，“大约”、“近似”或“大体上”表明所述数字允许有 $\pm 20\%$ 的变化。相应地，在一些实施例中，说明书和权利要求中使用的数值参数均为近似值，该近似值根据个别实施例所需特点可以发生改变。在一些实施例中，数值参数应考虑规定的有效数位并采用一般位数保留的方法。尽管本说明书一些实施例中用于确认其范围广度的数值域和参数为近似值，在具体实施例中，此类数值的设定在可行范围内尽可能精确。

[138]针对本说明书引用的每个专利、专利申请、专利申请公开物和其他材料，如文章、书籍、说明书、出版物、文档等，特此将其全部内容并入本说明书作为参考。与本说明书内容不一致或产生冲突的申请历史文件除外，对本说明书权利要求最广范围有限制的文件（当前或之后附加于本说明书中的）也除外。需要说明的是，如果本说明书附属材料中的描述、定义、和/或术语的使用与本说明书所述内容有不一致或冲突的地方，以本说明书的描述、定义和/或术语的使用为准。

[139]最后，应当理解的是，本说明书中所述实施例仅用以说明本说明书实施例的原则。其他的变形也可能属于本说明书的范围。因此，作为示例而非限制，本说明书实施例的替代配置可视为与本说明书的教导一致。相应地，本说明书的实施例不仅限于本说明书明确介绍和描述的实施例。

### 权利要求书

1、一种知识图谱数据融合方法，包括：

获取目标实体字段以及目标关系描述；所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据；其中，知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述；

确定一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子；

从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱。

2、如权利要求 1 所述的方法，所述目标实体字段以及所述目标关系描述由用户指定。

3、如权利要求 1 所述的方法，所述图谱算子由用户提供或自行生成。

4、如权利要求 1 所述的方法，还包括：基于所述目标实体字段、所述目标关系描述以及所述图谱算子获取共享知识图谱的本体定义数据，以及将所述融合知识图谱的本体定义数据通过知识图谱视图的形式表达。

5、如权利要求 1 所述的方法，所述实体字段对应有一个或多个属性字段，所述图谱算子用于实现以下操作中的一种或多种：

将目标实体字段对应的属性字段的实例值进行表达标准化处理；

将两个或更多个的目标实体字段进行融合，以得到融合实体字段；融合实体字段对应的属性字段来自所述两个或更多个的目标实体字段中的至少一个对应的属性字段；融合实体字段相关的关系描述包括所述两个或更多个的目标实体字段中的每一个相关的目标关系描述；

基于两个目标实体字段中的至少一个对应的属性字段，建立相应两个目标实体间的关系描述；

以及，调用自然语言处理模型确定数据实例中的相似实例，以便将数据实例中的相似实例进行融合。

6、如权利要求 1 所述的方法，所述从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱，包括：

确定图谱算子涉及的目标实体字段以及目标关系描述，作为最小子图的实体字段和关系描述；

从各知识图谱中获取最小子图的实体字段和关系描述对应的数据实例；

通过图谱算子处理最小子图的实体字段和关系描述对应的数据实例，得到最小子图；

从各知识图谱中获取最小子图的实体字段和关系描述以外的目标实体字段以及目标关系描述对应的数据实例，得到融合知识图谱除最小子图以外的子图。

7、如权利要求 1 所述的方法，所述从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例，并通过所述图谱算子处理所述数据实例以生成融合知识图谱在可信环境中执行。

8、如权利要求 7 所述的方法，还包括在所述可信环境中执行：

通过目标任务算法处理所述融合知识图谱,得到目标任务结果并输出;所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

9、如权利要求 8 所述的方法,所述目标任务算法由用户指定。

10、如权利要求 1 所述的方法,所述两个或更多个知识图谱来自一个或多个知识图谱提供方。

11、一种知识图谱数据融合系统,包括:

目标数据获取模块,用于获取目标实体字段以及目标关系描述;所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据;其中,知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述;

图谱算子确定模块,用于确定一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子;

融合图谱生成模块,用于从所述两个或更多个知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例,并通过所述图谱算子处理所述数据实例以生成融合知识图谱。

12、一种知识图谱数据融合装置,包括至少一个存储介质和至少一个处理器,所述至少一个存储介质用于存储计算机指令;所述至少一个处理器用于执行所述计算机指令以实现如权利要求 1-10 中任一项所述的知识图谱数据融合方法。

13、一种知识图谱数据处理方法,包括:

向服务方指定目标实体字段以及目标关系描述;所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据;其中,知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述;

从所述服务方处获取融合知识图谱和/或从所述服务方处获取目标任务结果;所述融合知识图谱通过图谱算子处理数据实例生成,所述数据实例基于所述目标实体字段以及所述目标关系描述从所述两个或更多个知识图谱中获取;所述目标任务结果通过目标任务算法处理所述融合知识图谱得到;所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

14、如权利要求 13 所述的方法,还包括:

生成一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子,并发送给所述服务方。

15、如权利要求 13 所述的方法,还包括:

向所述服务方指定目标任务算法。

16、如权利要求 13 所述的方法,还包括:

从所述服务方获取通过知识图谱视图的形式表达的融合知识图谱的本体定义数据;所述融合知识图谱的本体定义数据基于所述目标实体字段、所述目标关系描述以及所述图谱算子获取。

17、一种知识图谱数据处理系统,包括:

目标数据指定模块,用于向服务方指定目标实体字段以及目标关系描述;所述目标实体字段以及目标关系描述选自两个或更多个知识图谱的本体定义数据;其中,知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述;

结果获取模块,用于从所述服务方处获取融合知识图谱和/或从所述服务方处获取目标任务结果;所述融合知识图谱通过图谱算子处理数据实例生成,所述数据实例基于所述目标实体字段以及所述目标关系描述从所述两个或更多个知识图谱中获取;所述目标任务结果通过目标任务算法处理所述融合知识图谱得到;所述目标任务算法包括图谱规则推理算法或者基于图谱的机器学习模型预测算法。

18、一种知识图谱数据处理装置,包括至少一个存储介质和至少一个处理器,所述至少一个存储介质用于存储计算机指令;所述至少一个处理器用于执行所述计算机指令以实现如权利要求 13-16 中任一项所述的知识图谱数据融合方法。

19、一种知识图谱数据融合方法,包括:

获取两个或更多个知识图谱的本体定义数据;其中,知识图谱的本体定义数据包括用于定义实体的实体字段以及用于定义实体间关系的关系描述;

分别从各知识图谱的本体定义数据中筛选出目标实体字段以及目标关系描述,确定一个或多个用于对所述目标实体字段以及所述目标关系描述进行融合处理的图谱算子,进而得到融合知识图谱的本体定义数据;

从各知识图谱中获取所述目标实体字段以及所述目标关系描述对应的数据实例,并通过所述图谱算子处理所述数据实例以生成融合知识图谱。