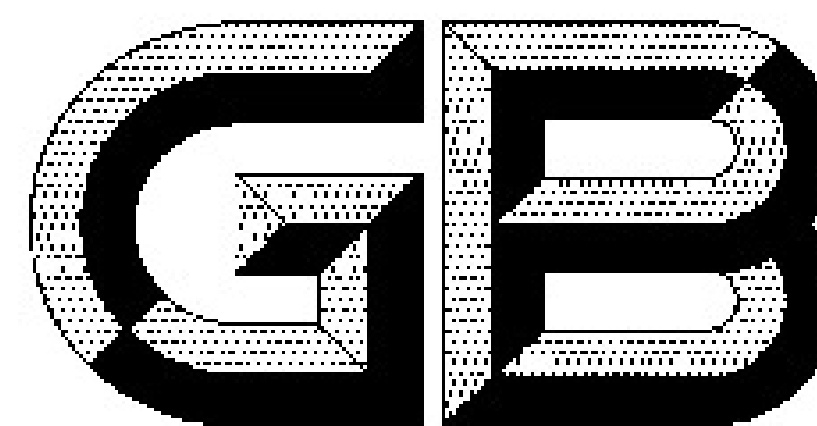




中华人民共和国国家标准

GB/T 42131—2022

人工智能 知识图谱技术框架

Artificial intelligence—Technical framework of knowledge graph

2022-12-30 发布

2023-07-01 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 3

5 概述 3

 5.1 知识图谱概念模型 3

 5.2 知识图谱技术框架 4

6 知识图谱供应方 6

 6.1 知识图谱供应方的输入 6

 6.2 知识图谱供应方的输出 6

 6.3 知识图谱供应方的主要活动 6

 6.3.1 活动流程 6

 6.3.2 知识表示 7

 6.3.3 知识建模 9

 6.3.4 知识获取 10

 6.3.5 知识融合 13

 6.3.6 知识存储 14

 6.3.7 知识计算 16

 6.3.8 知识溯源 17

 6.3.9 知识演化 18

 6.3.10 质量保障 19

7 知识图谱集成方 19

 7.1 知识图谱集成方的输入 19

 7.2 知识图谱集成方的输出 19

 7.3 知识图谱应用系统主要构成 20

 7.4 知识图谱集成方的主要活动 20

 7.4.1 活动流程 20

 7.4.2 需求分析 21

 7.4.3 系统设计 22

 7.4.4 知识图谱集成 23

 7.4.5 知识图谱应用系统开发 24

 7.4.6 系统维护 25

7.4.7 质量保障 26

8 知识图谱用户..... 26

8.1 知识使用者 26

8.1.1 知识使用者的输入 26

8.1.2 知识使用者的输出 27

8.1.3 主要活动 27

8.2 知识维护者 27

8.2.1 知识维护者的输入 27

8.2.2 知识维护的输出 27

8.2.3 主要活动 27

8.3 知识提供者 27

8.3.1 知识提供者的输入 27

8.3.2 知识提供者的输出 27

8.3.3 主要活动 27

9 知识图谱生态合作伙伴..... 28

9.1 知识图谱生态合作伙伴的输入 28

9.2 知识图谱生态合作伙伴的输出 28

9.3 主要活动 28

附录 A（资料性） 知识图谱生态合作伙伴子角色说明 29

参考文献 30

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：中国电子技术标准化研究院、江苏赛西科技发展有限公司、深圳赛西信息技术有限公司、清华大学、成都数联铭品科技有限公司、中电科大数据研究院有限公司、沈阳东软智能医疗科技研究院有限公司、北京华宇元典信息服务有限公司、北京国双科技有限公司、青岛百洋智能科技股份有限公司、网智天元科技集团股份有限公司、华为云计算有限公司、北京百分点科技集团股份有限公司、四川大学、上海依图网络科技有限公司、腾讯云计算(北京)有限责任公司、天津大学、中国医学科学院生物医学工程研究所、同方知网数字出版技术股份有限公司、北京车之家信息技术有限公司、北京京航计算通讯研究所、南华大学、绿盟科技集团股份有限公司、海信集团控股股份有限公司、浙商银行股份有限公司、北京海致星图科技有限公司、之江实验室、南京行者易智能交通科技有限公司、山东省人工智能研究院、东软集团股份有限公司、山东亿云信息技术有限公司、华为技术有限公司、山东省计算中心(国家超级计算济南中心)、电科云(北京)科技有限公司、南京视察者智能科技有限公司、中国电信股份有限公司研究院、北京智通云联科技有限公司、厦门邑通软件科技有限公司、上海智能制造功能平台有限公司、上海智能制造系统创新中心有限公司、南京柯基数据科技有限公司、南京航空航天大学、中国电子科技集团公司第二十八研究所、上海工程技术大学、中电莱斯信息系统有限公司、北京富通东方科技有限公司、云从科技集团股份有限公司、南瑞集团有限公司(国网电力科学研究院有限公司)、杭州海康威视数字技术股份有限公司、北京百度网讯科技有限公司、中科软科技股份有限公司、海义知信息科技(南京)有限公司、上海交通大学、厦门盈趣科技股份有限公司、厦门渊亭信息科技有限公司、北京智谱华章科技有限公司、富士康工业互联网股份有限公司、广州广电运通金融电子股份有限公司、北京电解智科技有限公司、浙江柏视医疗科技有限公司、如你所视(北京)科技有限公司、云孚科技(北京)有限公司、暗链科技(深圳)有限公司、中国电力科学研究院有限公司、浙江创邻科技有限公司、京东方科技集团股份有限公司、中电科新型智慧城市研究院有限公司、北京文因互联科技有限公司、云南昆船设计研究院有限公司、海尔优家智能科技(北京)有限公司、达而观信息科技(上海)有限公司、深圳市矽赫科技有限公司、郑州南河星科技有限公司、青岛海容商用冷链股份有限公司、医渡云(北京)技术有限公司。

本文件主要起草人：李瑞琪、李佳、韩丽、郭楠、贾仕齐、韦莎、李涓子、王伟光、查琳、郭晓妮、詹青、杨帆、贾承斌、侯磊、程序、李琴、邹丽华、杨斌、陈琦、苏海波、周磊、王海涛、石智中、白洋、蒲江波、刘丁泉、袁军、刘永彬、杨雷、宋明艳、谭培波、高雪松、陈嘉俊、熊林海、钟礼斌、何舟、李军、尹青云、陈通、高永超、陆保国、谢泽宇、张超、刘煜、柳军、王飞、王虎斌、郑毅、黄子珍、陈溪、刘安安、王鑫、张清易、舒明雷、单珂、周福辉、方志军、吴鹏亮、钟臻哲、宋勋超、王洋、陈建成、洪万福、张鹏、严红、孙佩霞、胡芳槐、吴文旷、艾青、石晓东、赵耕弘、高艳、王秋月、李瑞瑞、陈光、段飞虎、王博宇、颜子夜、孙林、曾俊瑀、李爽、陈佳林、吴刚、李军、欧阳纯萍、王芳杰、聂为之、赵春昊、傅洛伊、谈元鹏、曲珊、周研、肖向春、刘博、李卫东、李源、张娟、宋文韬、张学琴、苏婧仪、李钊、蔡巍、李东海、丁军、吴刚、王先庆、张文斌、徐佳吉、姜伟浩、姚博、蔺静茹、苏玉召、马力、刘海涛、符海芳、鲍捷、张明英、张浩男、李笑如、王启迪、金耀辉、姜伟浩、王文广、王艺杰、孙伟、张树刚、洪鹏达、万亚平、蒋炜、张楠、郎俊奇、焦飞、马洪奎、杨娟、胡芳槐、李论、曹扬、张晨、李林峰。

人工智能 知识图谱技术框架

1 范围

本文件给出了知识图谱的概念模型和技术框架,规定了知识图谱供应方、知识图谱集成方、知识图谱用户、知识图谱生态合作伙伴的输入、输出、主要活动和质量一般性能等要求。

本文件适用于知识图谱及其应用系统的构建、应用、实施与维护。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

知识 knowledge

通过学习、实践或探索所获得的认识、判断或技能。

[来源:GB/T 23703.2—2010, 2.1]

3.2

实体 entity

独立存在的对象。

3.3

实体类型 entity type

一组具有相同属性的实体集合的抽象。

3.4

知识元素 knowledge element

描述某一事物或概念的不必再分且独立的知识单位。

注:本文件中谈及的实体、实体类型(概念)、属性、关系、关系类型、事件、规则等统称为知识元素。

3.5

联系 association

两个或多个对象间语义上的关联。

[来源:ISO 13374-2:2007, B.2.2.4, 有修改]

3.6

知识图谱 knowledge graph

以结构化形式描述的知识元素及其联系的集合。

3.7

知识单元 knowledge unit

按照一定关系组织的一组知识元素的集合。

3.8

本体 ontology

表示实体类型以及实体类型之间关系、实体类型属性类型及其之间关联的一种模型。

注：又称本体模型。

[来源：ISO/IEC 21838-1:2021, 3.14, 有修改]

3.9

图式 schema

本体模型的规范化表达。

[来源：GB/T 5271.28—2001, 28.02.07, 有修改]

3.10

属性 attribute

一类对象中所有成员公共的特征。

[来源：GB/T 40216—2021, 3.1.2]

3.11

关系 relation

实体、实体类型、实体组合或实体类型组合间的联系。

注：关系可描述实体类型和实体类型、实体类型和实体、实体和实体之间的关联方式。

3.12

事件 event

发生在某个特定时间点或时间段、某个特定地域范围内,由一个或者多个角色参与的一个或者多个动作组成的事情或者状态的改变。

注：可表示为具有时间属性的实体和关系的组合。

3.13

实例 instance

某个实体类型或关系类型的具体范例。

3.14

知识图谱供应方 knowledge graph supplier

使用数据、知识等构建知识图谱以满足特定需求,并提供基于知识图谱的基础工具或服务的组织。

注：基础工具和服务指可基于其构建复杂的应用程序或系统的中间件。

3.15

知识图谱集成方 knowledge graph integrator

根据知识应用需求,将知识图谱、信息系统或服务进行整合,提供知识图谱应用系统及服务的组织。

3.16

知识图谱用户 knowledge graph user

使用知识图谱应用系统及配套服务支持以满足自身需要的组织或个人。

注：知识图谱用户可对外输出必要数据或知识。

3.17

知识图谱生态系统合作伙伴 knowledge graph ecosystem partner

为知识图谱供应方、集成方和用户提供知识图谱构建和应用所必需的信息基础设施、数据、工具、方法、标准和机制等的组织。

3.18

知识表示 knowledge representation

利用机器能够识别和处理的符号和方法描述人类的知识的活动。

3.19

知识建模 knowledge modeling

构建知识图谱的本体及其形式化表达的活动。

注：知识建模活动可包括实体类型定义、关系定义及属性定义。

3.20

知识获取 knowledge acquisition

从不同来源和结构的输入数据中提取知识的活动。

注：知识获取的数据源通常按数据组织结构的维度可分为结构化数据、半结构化数据、非结构化数据（如纯文本、音频和视频数据等）。

3.21

知识融合 knowledge fusion

整合和集成知识单元（集），并形成拥有全局统一知识标识的知识图谱的活动。

3.22

知识存储 knowledge storage

设计存储架构，并利用软硬件等基础设施对知识进行存储、查询、维护和管理的活动。

注：常见的知识存储方式分为：基于关系数据库的存储方式、基于图数据库的存储方式、基于资源描述框架（RDF）数据库的存储方式等。

3.23

知识计算 knowledge computing

基于已构建的知识图谱和算法，发现/获得隐含知识并对外提供知识服务能力的活动。

注：知识计算可分为统计分析、推理计算等。知识的统计分析是对知识图谱蕴含知识结构及其特征的统计与归纳；知识的推理计算是从已有的事实或关系推断出实现知识图谱隐性知识的发现与挖掘。

3.24

知识溯源 knowledge provenance

在知识图谱全生存周期中追踪原始数据向知识转化的活动。

3.25

知识演化 knowledge evolution

随本体模型、数据资源等变化产生的新知识对原有知识的补充、更新或重组的活动。

注：通过知识计算得出的补全知识也可触发知识演化。

4 缩略语

下列缩略语适用于本文件。

API：应用程序编程接口（Application Programming Interface）

RDF：资源描述框架（Resource Description Framework）

SDK：软件开发工具包（Software Development Kit）

5 概述

5.1 知识图谱概念模型

知识图谱的概念模型可划分为本体层和实例层，如图 1 所示。其中，本体层由实体类型和其属性、实体类型间关系类型、规则等本体相关知识元素构成；实例层是对本体层的实例化，由实体类型对应的实体及其属性以及实体间关系等实体相关知识元素构成。

图 1 示出的知识图谱概念模式的主体是实体。实体是真实对象的抽象,实体类型是某类实体的进一步抽象。基于不同层次的抽象,图中的本体层与实例层是相对的。构建某个知识领域的某个层次的特定知识图谱时,“实体”这个抽象称呼将使用所关注的特定对象的具体名称取代。图中名为“属性”的两个方框是分别针对本体层的所有实体类型和实例层的所有实体。本体层的属性是指对应实体类型的属性,各个属性是概括性描述;实例层的属性是指对应实体的属性,是某实体类型实例的属性的具体描述。同时,多个实体和关系的组合可以构成新的复杂实体,例如:由时间、人物、地点等要素构成的事件,由不同模块构成的产品等。

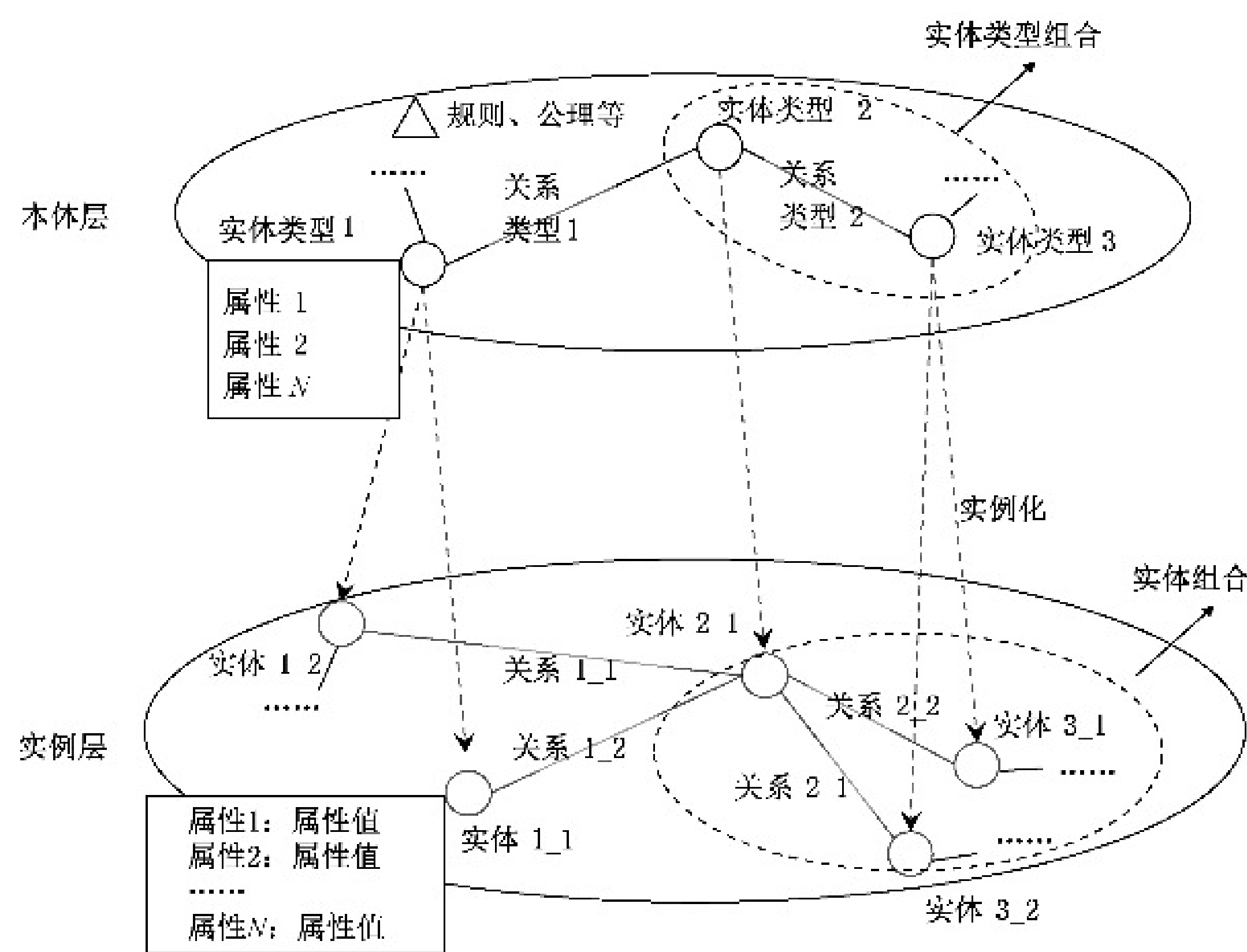


图 1 知识图谱的概念模型

5.2 知识图谱技术框架

图 2 示出从构建到使用知识图谱涉及的各类利益相关方和各类技术活动的技术框架,技术活动包括:知识图谱的构建、基于知识图谱的产品或服务的开发、知识图谱的使用以及知识图谱开发和使用的支持等四类。

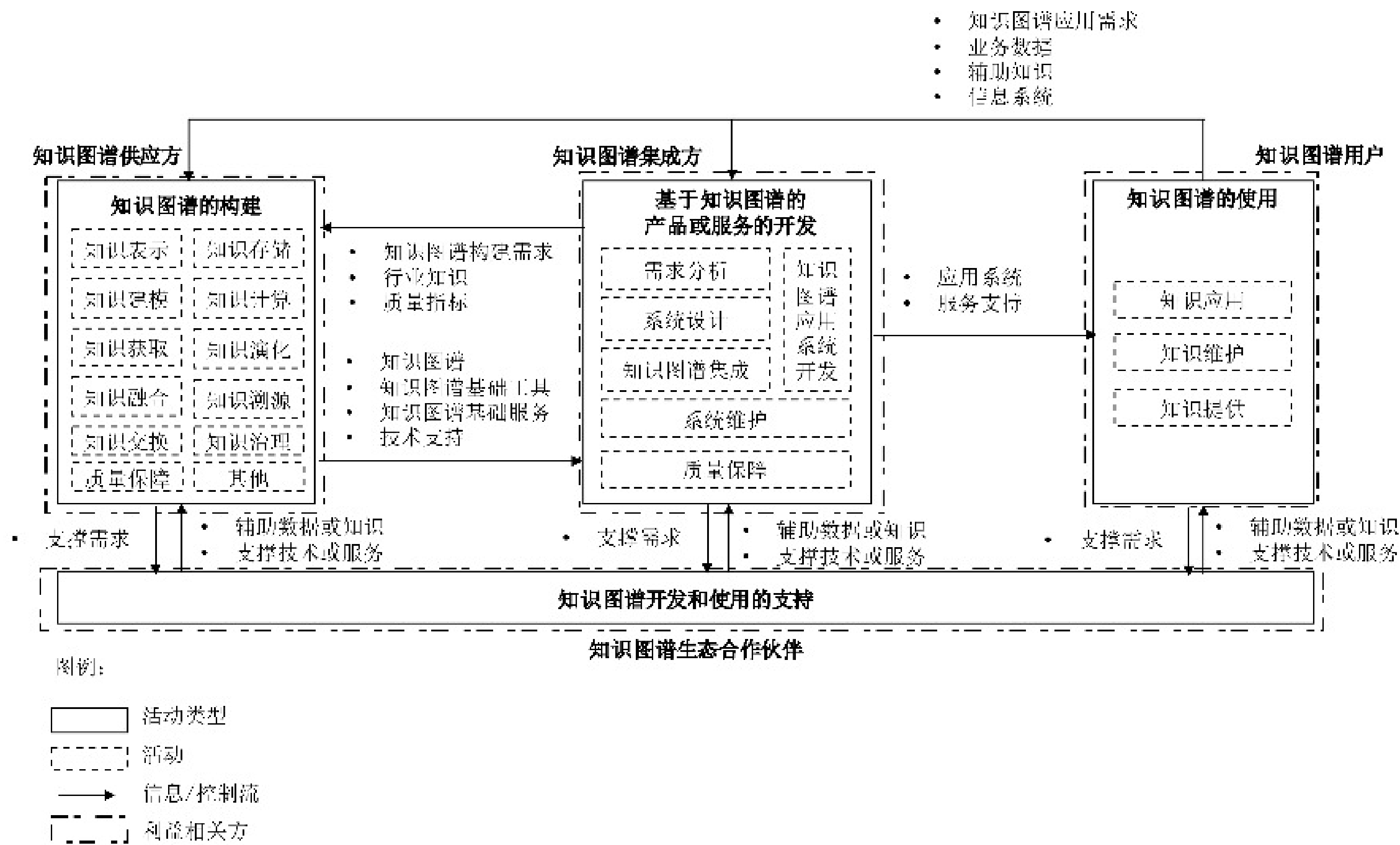


图 2 知识图谱技术框架

图 2 示出的四类知识图谱相关活动简述如下。

- a) 知识图谱的构建：此组活动主要包括知识表示、知识建模、知识获取等活动。其主要目标是构建出所需的知识图谱，同时开发出相应的基础工具和/或服务。此组活动的主要依据是知识图谱应用需求和质量要求；往往需要行业知识、业务数据、辅助知识等予以支持。
- b) 基于知识图谱的产品或服务开发：此组活动主要包括需求分析、系统设计、知识图谱集成等活动。这些活动的执行基于上述 a) 描述的活动构建的知识图谱和相应的知识图谱应用需求等完成知识图谱应用系统的开发和集成，并提供配套的产品或服务。
- c) 知识图谱的使用：此组活动主要包括知识应用、知识维护、知识提供等活动。这些活动的执行基于上述 b) 描述的活动产生的知识图谱应用系统或服务。通过这些活动完成知识的使用和维护，并对外提供必要的知识。
- d) 知识图谱开发和使用的支持：此组活动主要包括基础设施提供、数据提供、安全保障、咨询评估等。它们对上述 a)、b) 和 c) 描述的活动的执行提供必要支持，例如：提供辅助数据或知识、支撑技术或服务。

上述四类活动分别主要由以下四类参与者执行：

- a) 知识图谱供应方：主要执行知识图谱的构建和提供活动；
 - b) 知识图谱集成方：主要执行基于知识图谱的工具或服务开发和集成活动；
 - c) 知识图谱用户：主要执行知识图谱的使用活动；
 - d) 知识图谱生态系统合作伙伴：主要执行知识图谱开发和使用的支持活动。
- 每类参与者有其主要执行的活动，同时可能执行涉及四类活动中的多项活动。

6 知识图谱供应方

6.1 知识图谱供应方的输入

知识图谱供应方的输入包括但不限于：

- a) 知识图谱应用需求：知识图谱用户对应用知识图谱提出的要求，如业务需求、应用约束、数据现状等；
- b) 知识图谱构建需求：知识图谱集成方对构建知识图谱提出的要求，如知识范围、知识粒度、知识图谱规模等；
- c) 业务数据：构建知识图谱所需的基础数据、行业数据和其他必要数据；
- d) 辅助知识：指导或支撑知识图谱构建的常识、行业知识、专家知识等；
- e) 支撑技术和服务：构建知识图谱所需的技术和服务支持，如数据预处理、数据标注工具等；
- f) 质量指标：知识图谱用户或知识图谱集成方提出的知识图谱质量要求和性能指标。

6.2 知识图谱供应方的输出

知识图谱供应方的输出包括但不限于：

- a) 知识图谱；
- b) 知识图谱基础工具：具有知识获取、检索、关系推理、可视化和维护等基本功能的知识图谱工具或系统等；

注 1：部分知识图谱工具也具备知识管理相关基础功能。

- c) 知识图谱基础服务：基于知识图谱以接口等形式提供知识查询或计算结果调用的服务；

注 2：知识图谱供应方出于安全、所有权等考虑会提供基础服务；知识图谱集成方部分情况下通过使用知识图谱基础服务在其上层开发和集成新的产品或服务。

- d) 技术支持：提供知识图谱基础工具或基础服务在集成过程中所需的技术支持。

6.3 知识图谱供应方的主要活动

6.3.1 活动流程

知识图谱供应方的主要活动流程见图 3，可包括：

- a) 知识表示：利用机器能够识别和处理的符号和方法描述人类的知识的活动；
- b) 知识建模：构建知识图谱的本体及其形式化表达的活动；
- c) 知识获取：从不同来源和结构的输入数据中提取知识的活动；
- d) 知识融合：整合和集成知识单元(集)，并形成拥有全局统一知识标识的知识图谱的活动；
- e) 知识存储：设计存储架构，并利用软硬件等基础设施对知识进行存储、查询、维护和管理的活动；
- f) 知识计算：基于已构建的知识图谱和算法，发现/获得隐含知识并对外提供知识服务能力的活动；
- g) 知识演化：随本体模型、数据资源等变化产生的新知识对原有知识的补充、更新或重组的活动；
- h) 知识溯源：在知识图谱全生存周期中追踪原始数据向知识转化的活动；
- i) 质量保障：保障知识内容、知识图谱构建过程、知识图谱基础服务或工具的质量的活动。

除上述主要活动，知识图谱供应方还可包括知识交换、知识治理等其他可选活动。

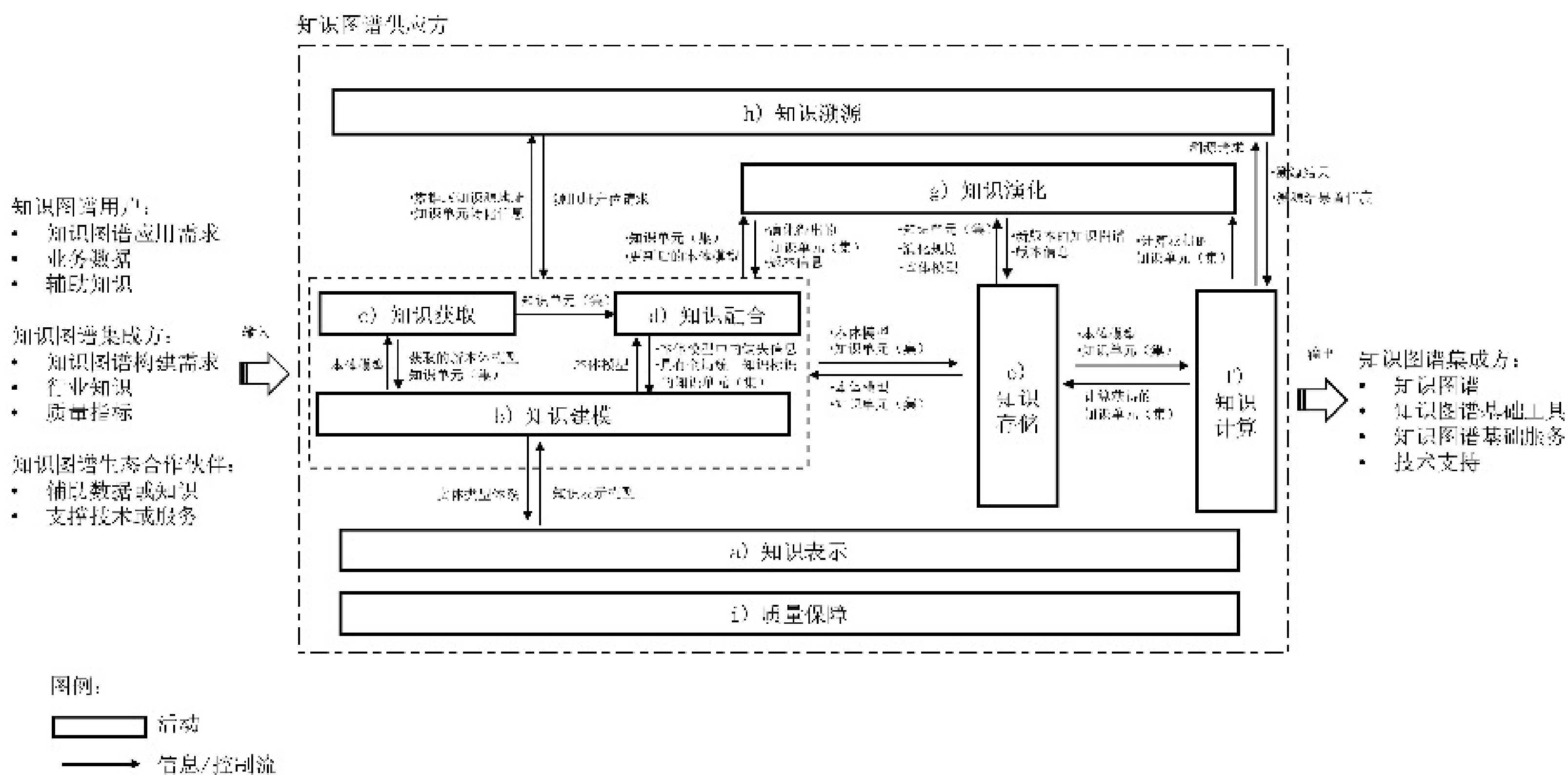


图 3 知识图谱供应方主要活动流程图

6.3.2 知识表示

6.3.2.1 活动输入

知识表示活动的输入包括但不限于如下内容。

- a) 知识图谱应用需求，如：
 - 1) 业务需求：拟解决的业务问题及拟达成的业务目标；
 - 2) 应用场景：拟部署应用的具体业务场景；
 - 3) 应用约束：知识图谱应用过程中应遵循的相关要求、标准、法律法规等；
 - 4) 知识背景：知识表示专家具有的学科背景、技术背景、领域背景等；
 - 5) 应用反馈：知识图谱供应方其他活动应用知识表示模型的意见及建议。
- b) 知识图谱构建需求。
- c) 数据需求：数据包括基础训练与测试数据、业务数据等，主要用于支持知识表示学习、知识获取等环节算法模型的设计、训练测试及后续知识图谱的构建。
- d) 质量指标。
- e) 实体类型体系。

6.3.2.2 活动输出

知识表示活动的输出包括但不限于如下内容。

- a) 知识表示模型，可包括：
 - 1) 知识表示框架：知识表示结构和具体表现形态；
 - 2) 知识表示元素：知识表示过程中需要使用的元素及其含义，如实体类型、实体、关系类型、推理规则等；
 - 3) 知识表示要求：知识表示过程中需要遵守的规则、约束；
 - 4) 知识表示适用范围：知识表示模型的边界、范围和限制。
- b) 知识表示模型质量评价体系。

注：知识表示模型通常可分为基于离散符号的知识表示和基于连续向量的数值知识表示。

6.3.2.3 任务组成

知识表示活动的任务组成见图 4,包括但不限于：

- a) 定义知识表示需求,如拟解决的业务问题、拟实现的业务目标等；
- b) 定义或确定拟遵循的规则、约束,如业务规则及相关约束等；

注 1：面向特殊领域,对不适于或缺失已有规则和约束的场景设计规则和约束。

- c) 定义或选择知识表示形式；

注 2：面向特殊领域,针对需求采用或设计多元组、框架等知识表示形式。

- d) 定义和序列化知识表示元素,并制定知识表示过程应遵循的相关约束、通用规则等；
- e) 定义知识表示模型适用范围,如适用场景、不适用的场景、使用的注意事项等；
- f) 定义知识表示模型评价体系；
- g) 评估知识表示能力。

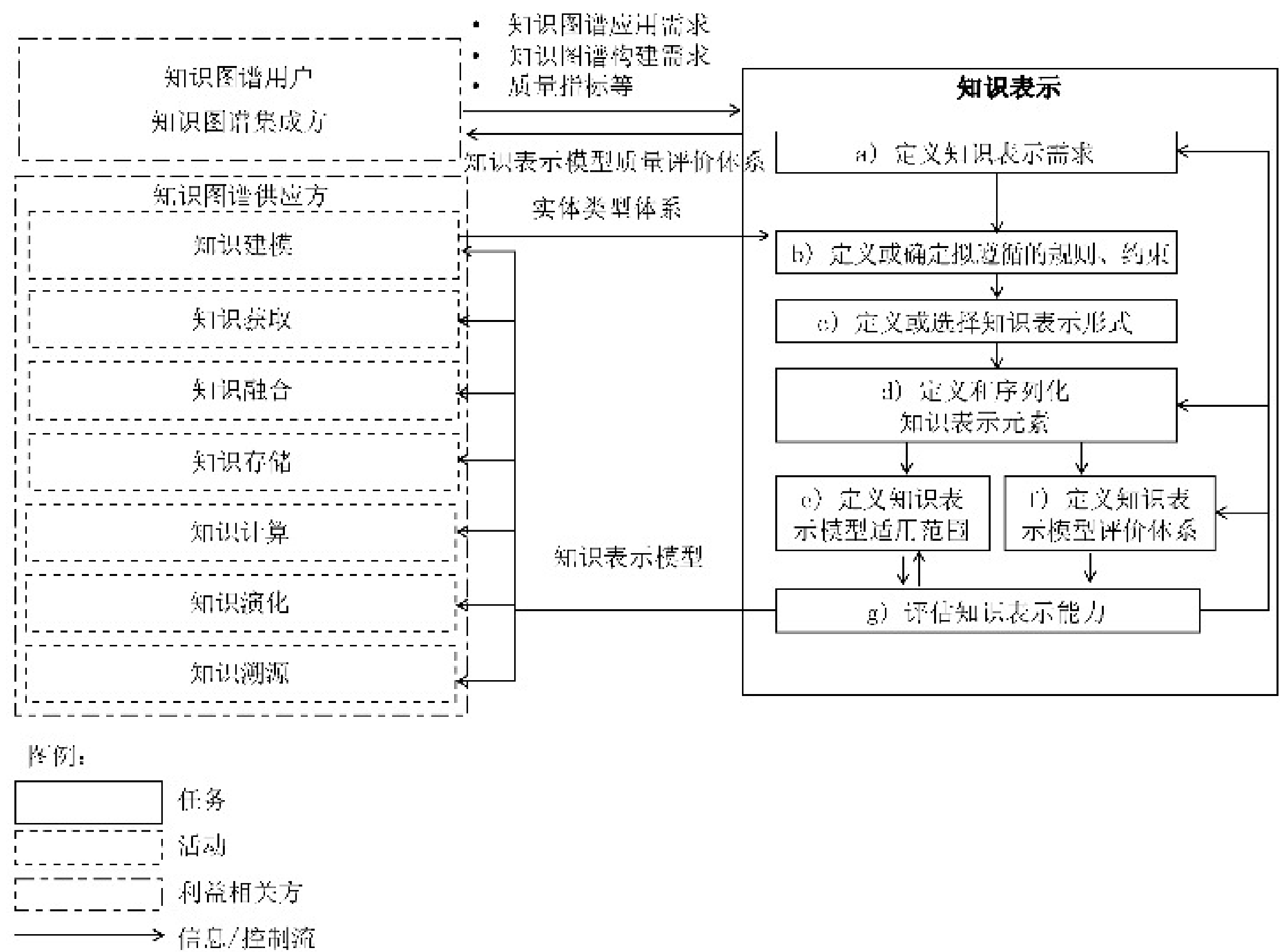


图 4 知识表示任务流程图

6.3.2.4 质量一般性能

用于描述知识表示活动质量的一般性能包括但不限于：

- a) 可表达性:形成的知识表示模型完整表达特定领域业务所需知识且可被实施人员理解的程度；
- b) 可实现性:形成的知识表示模型是否可被计算机识别及被算法实现；
- c) 严密性:形成的知识表示模型是否可描述形式化的语法、语义及相关推理规则；
- d) 可维护性:形成的知识表示模型是否可支持知识图谱构建完成后知识单元的维护和管理。

6.3.3 知识建模

6.3.3.1 活动输入

知识建模活动的输入包括但不限于如下内容。

- a) 知识图谱应用需求,如:
 - 1) 业务需求;
 - 2) 应用场景;
 - 3) 应用约束;
 - 4) 数据现状:拟解决的业务问题相关数据探查的结果,包括数据字典、数据质量、数据量、已处理过的数据结构等;
 - 5) 应用反馈:知识图谱供应方其他活动应用本体模型的建议等。
- b) 知识图谱构建需求。
- c) 辅助知识,如:
 - 1) 行业知识,可包括术语字典、术语体系、行业指南、行业标准、其他行业知识等;
 - 2) 专家知识。
- d) 质量指标。
- e) 知识表示模型。
- f) 知识表示活动、知识获取活动和知识融合活动输出的实体类型、关系类型等知识单元。

6.3.3.2 活动输出

知识建模活动的输出包括但不限于如下内容。

- a) 本体模型:
 - 1) 实体类型体系,如实体类型;实体类型间的上下位关系;
 - 2) 实体类型的属性,如属性字段的类型;实体类型的唯一标识属性;属性是否具有唯一性等;
 - 3) 实体类型间的关系,如关系是否有方向(有向关系和无向关系);关系是否有传递;是否为1对1关系。

注:事件可视为实体的一种,事件类型也可作为本体模型的一部分。

- b) 图式(schema)。

6.3.3.3 任务组成

知识建模活动的任务组成见图5,包括但不限于如下内容。

- a) 确定知识的领域和范畴。
- b) 确定现有可复用本体模型,可复用本体模型的确认原则可包括:
 - 1) 如非必要,不宜新增实体;
 - 2) 实体类别融合原则:如果两类实体类别的实例相同,可融合对应实体类别;
 - 3) 实体类别分拆原则:如果某实体类别的互斥属性较多,可拆成多个细分实体类别。
- c) 确定知识范畴内的关键术语。
- d) 构建实体类别层级体系。
- e) 定义实体类别的属性与关系。
- f) 定义应用需求相关的规则、公理等(可选)。
- g) 确定并创建本体模型及图式。
- h) 评估本体模型质量。

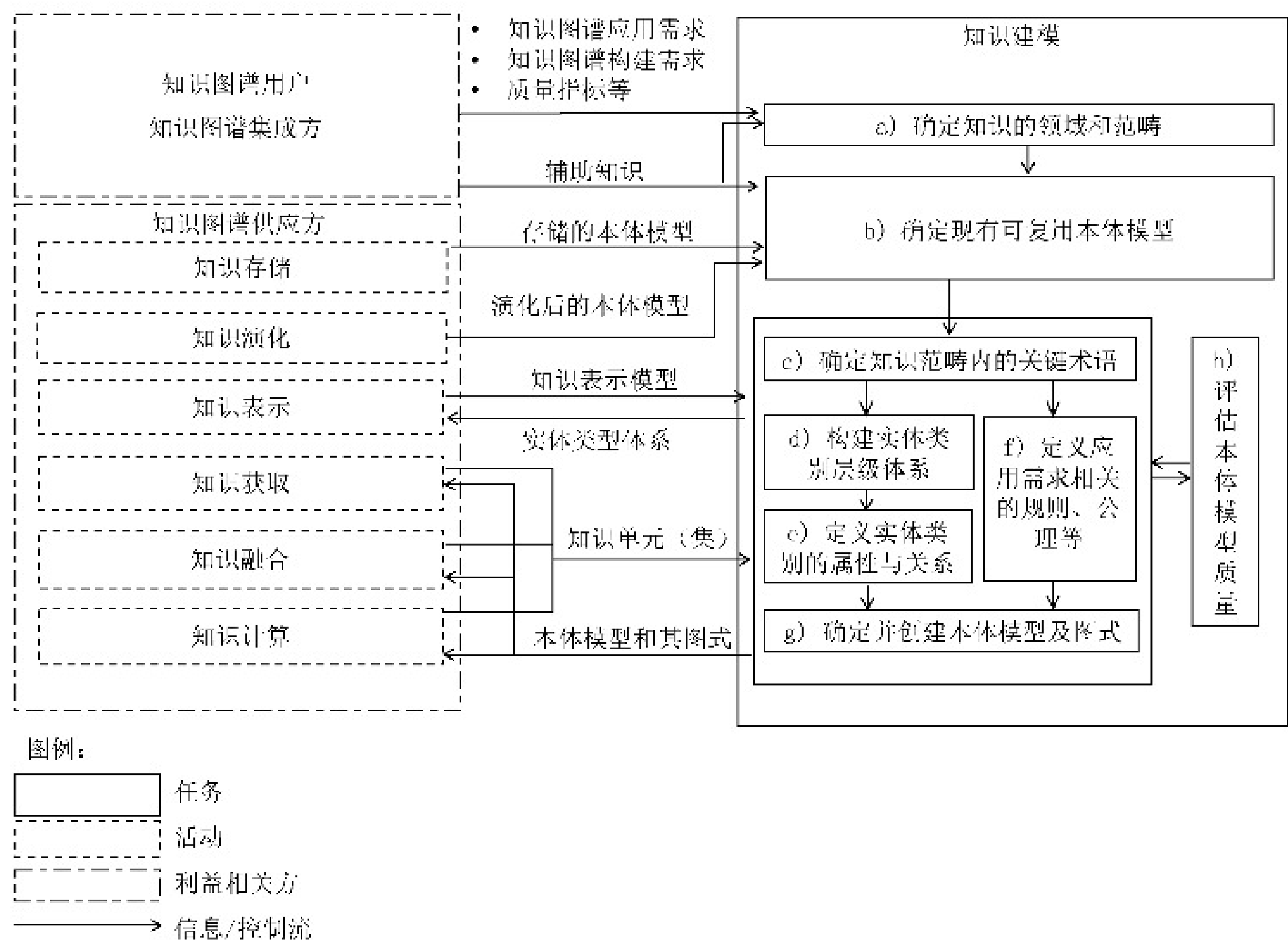


图 5 知识建模任务流程图

6.3.3.4 质量一般性能

知识建模活动宜形成一个定义良好的本体模型及其图式，并根据输入数据特征和应用特点进行知识管理以减少数据冗余并提高应用效率。用于描述知识建模活动质量的一般性能包括但不限于如下内容。

- a) 合理性：形成的本体模型中实体类型是否划分合理和实体类型间关系描述合理。
- b) 可用性：形成的本体模型是否可支持后续知识获取、知识融合、知识计算等活动。
- c) 完整性：形成的本体模型是否可支持事件和时间序列等复杂知识的表示，并可支持匿名实体的使用。
- d) 可扩展性：形成的本体模型是否可支持实体类型及其属性与关系的添加、删除和修改。
- e) 兼容性：形成的本体模型是否可以实现与已有本体模型的兼容或继承等，如：
 - 1) 形成的本体模型是否可支持与本领域其他标准本体模型的兼容；
 - 2) 形成的本体模型是否可支持已有实体类型体系的继承；
 - 3) 形成的本体模型是否可支持业务或输入数据源变更或更新后实体类型属性的调整。
- f) 可复用性：形成的本体模型是否可支持多知识图谱间的复用。
- g) 简洁性：创建的本体模型图式是否可符合最小冗余原则以满足知识融合和知识计算中本体模型的应用和可视化要求。

6.3.4 知识获取

6.3.4.1 活动输入

知识获取活动的输入包括但不限于如下内容。

- a) 数据，可包括：

- 1) 结构化数据,如存储于关系型数据库中的业务数据;
 - 2) 半结构化数据,如百科数据等;
 - 3) 非结构化数据,如文档、图片、视频、音频等。
- b) 已有的本体模型和图式。

6.3.4.2 活动输出

知识获取活动的输出为知识单元(集),包括但不限于如下内容。

- a) 实体信息;
- b) 实体间的关系信息;
- c) 实体的属性信息;
- d) 本体模型中缺失信息,如:
 - 1) 实体类型信息;
 - 2) 实体类型间的关系信息;
 - 3) 实体类型的属性信息。

注:本体模型中缺失信息将反馈至知识建模活动,并进一步优化本体模型。

6.3.4.3 任务组成

知识获取活动的任务组成见图 6,包括但不限于如下内容。

- a) 针对结构化数据提取,子任务可包括:
 - 1) 选择映射方式;
 - 2) 定义和配置映射规则。
- b) 针对半结构化数据提取,子任务可包括:
 - 1) 针对百科类半结构化数据,设计提取器实现知识单元的提取;
 - 2) 针对网页类半结构化数据,设计包装器实现知识单元的提取。
- c) 针对非结构化数据提取,子任务可包括:
 - 1) 基于业务需求和知识建模活动形成的本体模型等,选择或设计适合业务场景和数据类型的知识提取算法;
 - 2) 完成算法测试并实现知识单元的提取。
- d) 人工录入:可基于业务需求和知识建模活动形成的本体模型等以人工方式提取并录入知识单元。
- e) 三或多元组提取,子任务可包括:
 - 1) 提取实体类型:基于提取后的知识单元识别领域实体类型;
 - 2) 提取实体:基于提取后的知识单元获取通用或者特定领域的实体;
 - 3) 提取关系:基于提取后的知识单元获取实体类型间、实体间的语义关系;
 - 4) 提取属性:基于提取后的知识单元获取实体类型或实体中包含的描述特征;

注:实体、关系或属性可联合抽取,联合抽取不依赖实体抽取活动。

 - 5) 提取事件等:基于提取后的数据结构化地提取事件等信息,如事件发生的时间、地点、发生原因、参与者等。
- f) 知识获取质量评估。

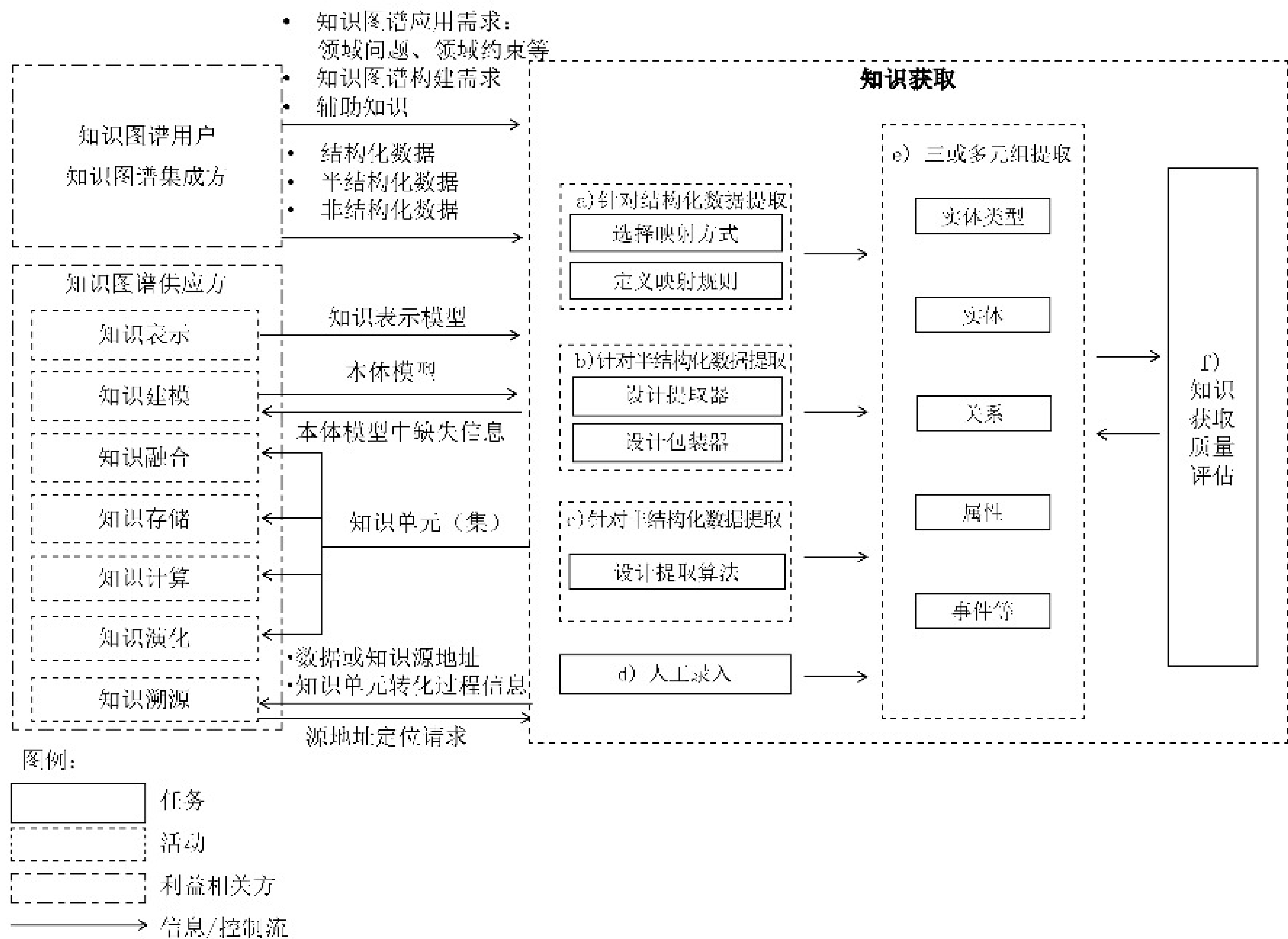


图 6 知识获取任务流程图

6.3.4.4 质量一般性能

知识获取活动质量的评估可通过获取知识单元与人工标注数据的比较得出，一般性能主要包括如下内容。

a) 精确率 $Precision_A$ ：度量已获取知识中正确知识的占比。计算公式见式(1)：

$$Precision_A = \frac{TP_A}{TP_A + FP_A} \dots\dots\dots (1)$$

式中：

TP_A ——真阳性：被识别并与真实相符的实体、关系或属性数量；

FP_A ——假阳性：被识别但与真实不符的实体、关系或属性数量。

b) 召回率 $Recall_A$ ：度量已获取知识对正确知识覆盖的程度。计算公式见式(2)。

$$Recall_A = \frac{TP_A}{TP_A + FN_A} \dots\dots\dots (2)$$

式中：

FN_A ——假阴性：被标注为真实但未识别的实体、关系或属性数量。

c) F1 测量值 $F1-score_A$ ：综合度量知识获取结果的精确性和完整性，计算公式见式(3)：

$$F1-score_A = 2 \times \frac{Precision_A \times Recall_A}{Precision_A + Recall_A} \dots\dots\dots (3)$$

d) 关系约束校验，如：

1) 是否为 1 对 1 关系（依据本体模型关系定义）；

- 2) 是否为 1 对 n 关系(依据本体模型关系定义);
- 3) 是否为多对多关系(依据本体模型关系定义)。
- e) 属性约束校验,如:
 - 1) 属性是否具有唯一性(依据本体模型属性定义);
 - 2) 属性是否要求非空(依据本体模型属性定义)。

注:对于知识获取的整体能力评估,可考虑实体-属性/关系-实体联合作为一组评测数据来进行精确率、召回率、F1 值的评估。

6.3.5 知识融合

6.3.5.1 活动输入

知识融合活动的输入包括但不限于:

- a) 知识建模活动输出的本体模型和图式;
- b) 知识获取活动输出的知识单元(集);
- c) 来自知识图谱外部的知识单元(集)。

6.3.5.2 活动输出

知识融合活动的输出包括但不限于如下内容。

- a) 统一的本体模型:
 - 1) 统一的实体类型;
 - 2) 统一的实体类型间的关系;
 - 3) 统一的实体类型属性。
- b) 具有全局统一知识标识且不冲突的知识单元。

注:识别的本体模型中缺失信息将反馈至知识建模活动,并进一步优化本体模型。

6.3.5.3 任务组成

知识融合活动的任务组成见图 7,包括但不限于如下内容。

- a) 本体对齐,子任务可包括:
 - 1) 实体类别对齐;
 - 2) 关系对齐:对齐不同本体模型中的等效关系,并对关系进行合并;
 - 3) 属性对齐:对齐不同本体模型中的等价属性,并合并为同一属性。
- b) 实体对齐:识别知识单元中等效的实体,可包括实体归一、实体消歧、属性值对齐与融合等。
- c) 知识一致性校验:校验融合后知识单元间的一致性。

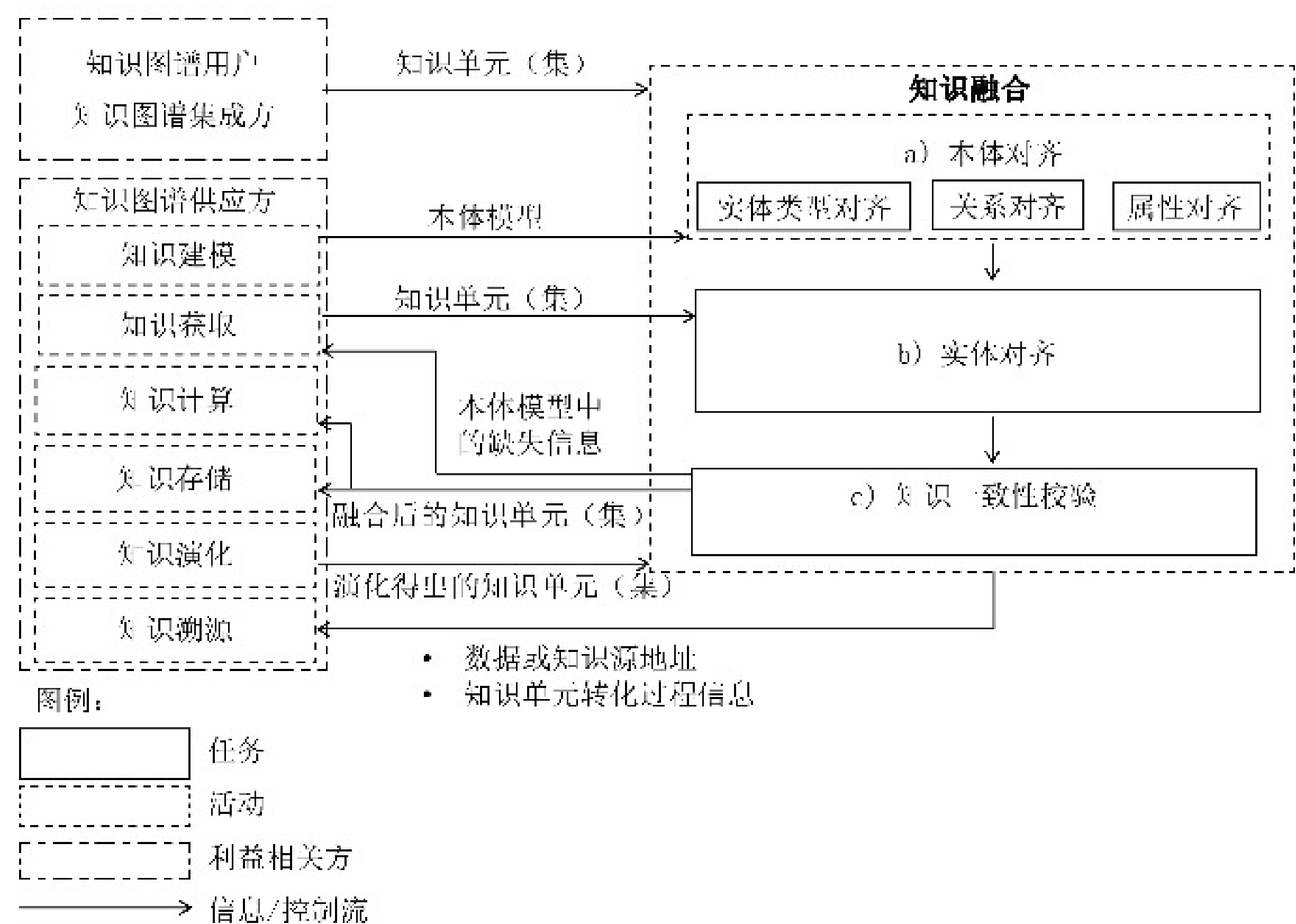


图 7 知识融合活动任务流图

6.3.5.4 质量一般性能

知识融合活动的质量可通过精确率、召回率、F1 测量值和前 N 命中率进行评估。

a) 精确率 Precision_F : 度量知识融合结果正确知识与总融合知识规模的比例, 计算公式见式(4)。

$$\text{Precision}_F = \frac{\text{TP}_F}{\text{TP}_F + \text{FP}_F} \dots\dots\dots (4)$$

式中：

TP_F ——真阳性: 融合结果中与真实相符的实体、关系或属性数量；
 FP_F ——假阳性: 融合结果中已融合但与真实不符的实体、关系或属性数量。

b) 召回率 Recall_F : 度量知识融合结果中已融合知识对正确知识覆盖的程度, 计算公式应满足式(5)。

$$\text{Recall}_F = \frac{\text{TP}_F}{\text{TP}_F + \text{FN}_F} \dots\dots\dots (5)$$

式中：

FN_F ——假阴性: 融合结果中待融合但未融合的实体、关系或属性数量。

c) F1 测量值 F1-score_F : 综合度量知识融合结果的精确性和完整性, 计算公式应满足式(6)。

$$\text{F1-score}_F = 2 \times \frac{\text{Precision}_F \times \text{Recall}_F}{\text{Precision}_F + \text{Recall}_F} \dots\dots\dots (6)$$

d) 前 N 命中率: 综合度量正确的实体、关系或属性排名在本活动计算得出的前 N 个待融合结果中的比例。

6.3.6 知识存储

6.3.6.1 活动输入

知识存储活动的输入包括但不限于：

a) 本体模型和其图式；

- b) 知识融合、知识计算、知识演化等活动输出的知识单元；
- c) 知识演化活动输出的时间版本信息；
- d) 知识表示模型；
- e) 知识建模、知识计算、知识演化、知识溯源等活动设计或输出的规则、约束、算法模型等。

6.3.6.2 活动输出

知识存储活动的输出包括但不限于如下内容。

- a) 设计或选型后的数据库系统,如:
 - 1) 知识查询和编辑接口；
 - 2) 可视化模块；
 - 3) 存储管理接口；

注 1: 知识存储管理接口,主要包括图式查询修改、副本数量查询修改、索引配置等。

 - 4) 支撑工具。

注 2: 支撑工具包括数据的导入导出工具、外部业务组件连接工具、性能监控工具、备份恢复工具等。
- b) 存储的知识单元。
- c) 存储的本体模型。
- d) 存储的规则、约束、算法模型等。

6.3.6.3 任务组成

知识存储活动的任务组成包括但不限于如下内容。

- a) 完成数据存储系统选型与设计,如:
 - 1) 根据对知识的存储和应用需求结合资源配置条件选择数据存储系统类型,如数据库、文件系统等；
 - 2) 基于选择的数据存储系统类型设计知识存储结构。
- b) 执行存储操作,如:
 - 1) 完成知识单元的存储；
 - 2) 完成知识单元的查询；
 - 3) 完成知识单元的维护,如知识单元的新增、删除、修改、更新等；
 - 4) 完成知识单元的可视化,如以图等可视化方式展现知识单元的内容。
- c) 完成存储管理,如:
 - 1) 完成数据库设置信息的配置；
 - 2) 记录知识存储过程的日志；
 - 3) 建立数据安全机制,保障数据存储安全。

6.3.6.4 质量一般性能

描述知识存储活动质量的一般性能包括但不限于：

- a) 效率:度量活动中时间占用、资源占用等程度；
- b) 可维护性:度量活动中知识单元维护与管理与数据库维护的能力；
- c) 安全性:度量活动中对知识单元进行保护以防止丢失或泄露的能力；
- d) 可用性:度量对知识图谱其他活动的支持能力,可包括知识单元导入导出、分布式存储、索引构建、全知识图谱检索等；
- e) 可靠性:度量是否建立知识单元的备份机制；
- f) 友好性:度量接入接口可支持的常见数据类型、界面友好程度及人机交互能力等。

6.3.7 知识计算

6.3.7.1 活动输入

知识计算活动的输入包括但不限于：

- a) 知识表示模型；
- b) 本体模型和其图式；
- c) 已存储的知识单元(集)；
- d) 基于知识图应用需求和构建需求获得的计算需求，如算法模型、(业务)规则、溯源请求、外部数据等；
- e) 从知识溯源活动的输入：溯源结果、溯源结果置信度等。

6.3.7.2 活动输出

知识计算活动的输出包括但不限于：

- a) 计算获得的知识单元(集)；
- b) 计算服务：为下游任务提供的知识图谱计算服务；

注：计算服务可包括图特征、子图、链接预测、图嵌入(Graph Embedding)、图搜索、算法模型、实体链接等。

- c) 输出至知识溯源活动：溯源请求。

6.3.7.3 任务组成

知识计算活动的任务组成见图 8，包括但不限于如下内容。

- a) 定义知识计算需求，如：
 - 1) 定义拟解决的业务问题和拟实现的计算目标；
 - 2) 分析依赖的本体模型、规则和约束。
- b) 设计计算所需的数据结构及算法模型。
- c) 执行知识计算流程并评估计算性能，如：
 - 1) 通过统计分析对知识图谱蕴含知识结构及其特征进行统计与归纳；
 - 2) 通过推理计算从已有的事实或关系中进行隐性知识的发现与挖掘。
- d) 基于挖掘的隐性知识补全缺失的知识单元。
- e) 通过接口等形式提供知识计算服务，如图基本信息统计、图搜索、三元组分类、链接预测等。

注：知识图谱应用过程中也可调用知识计算服务。

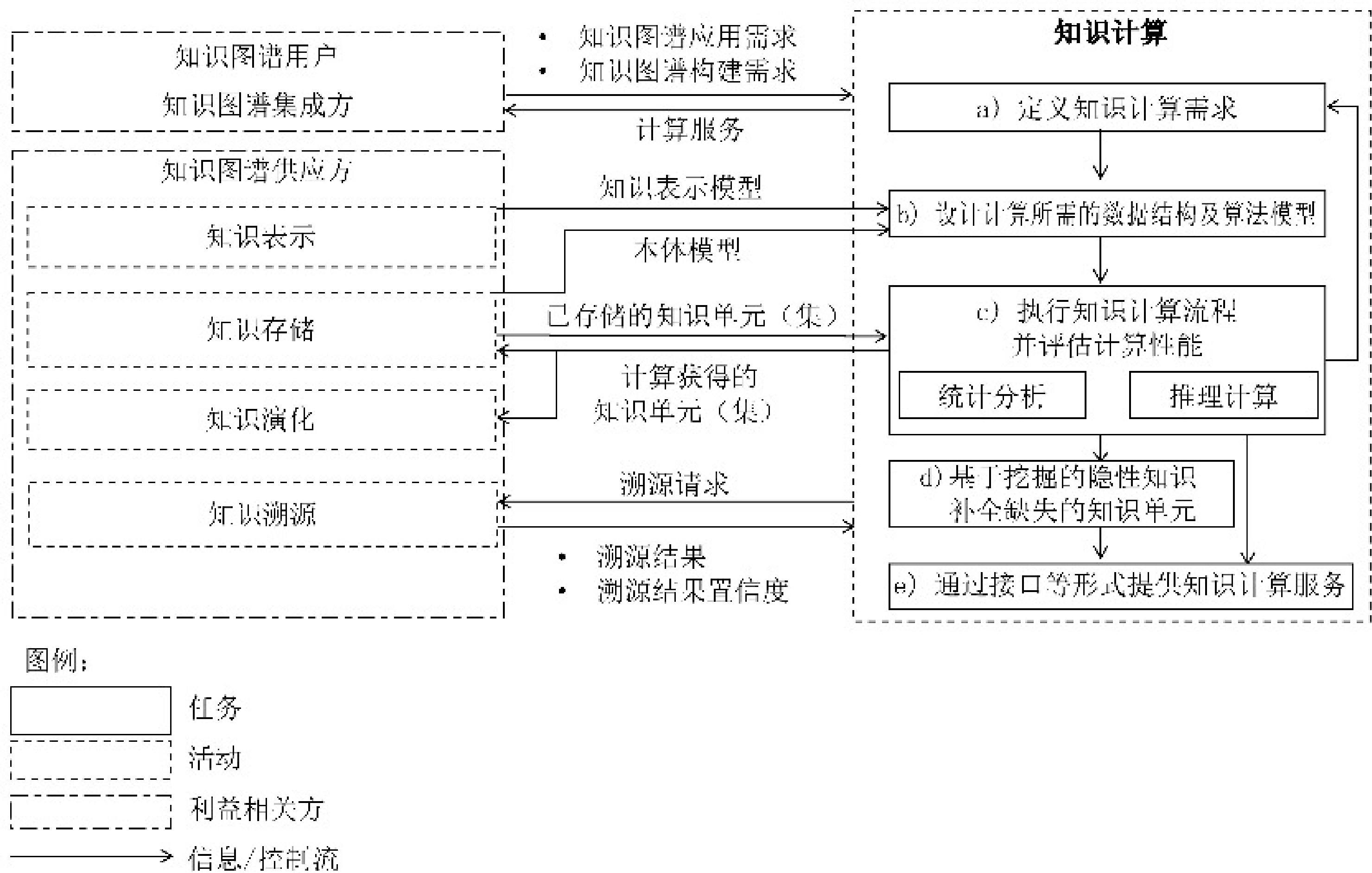


图 8 知识计算任务流程图

6.3.7.4 质量一般性能

描述知识计算活动质量的一般性能包括但不限于：

- a) 准确性：度量形成的算法模型在计算任务上的准确性；
- b) 性能：度量活动中资源占用、时间占用等情况；
- c) 资源消耗：度量活动中软硬件资源占用等情况；
- d) 响应时间：度量活动中时间占用情况；
- e) 计算能力：度量活动对算法、规则种类的支撑能力；
- f) 可解释性：度量计算过程的可解释程度，如活动中定义规则的逻辑可解释程度等；
- g) 可用性：度量计算服务的可访问程度。

6.3.8 知识溯源

6.3.8.1 活动输入

知识溯源活动的输入包括但不限于：

- a) 数据或知识源地址；
- b) 溯源请求；
- c) 知识建模、知识获取、知识融合、知识计算、知识演化等活动中知识单元转化的信息。

6.3.8.2 活动输出

知识溯源活动的输出包括但不限于如下内容。

- a) 源地址定位请求。
- b) 溯源结果，如：
 - 1) 知识源地址：用于标识知识来源的唯一资源访问地址；

- 2) 知识单元转化路径。
- c) 溯源结果置信度:用于评价溯源结果的可信度。

6.3.8.3 任务组成

知识溯源活动的任务组成包括但不限于:

- a) 定义知识源地址的结构化描述,如创建者或发布者、时间戳、版本、格式等;

注:时间戳指数据的时间维度和时间粒度,时间维度包括基于版本管理角度的时间维度和基于事实的时间维度。

- b) 设计知识溯源方案,如溯源模型、溯源方法的组织与表达等;
- c) 追溯知识来源,并完成知识定位;
- d) 确定置信度计算维度并完成溯源结果计算,如数据源的数量、各数据源的生成时间戳及权重等;
- e) 对原始数据和知识的更新进行版本管理。

6.3.8.4 质量一般性能

描述知识溯源活动质量的一般性能包括但不限于:

- a) 准确性:度量查询结果与目标知识的相关性;
- b) 完整性:度量查询结果与目标相关知识的比例;
- c) 响应时间。

6.3.9 知识演化

6.3.9.1 活动输入

知识演化活动的输入包括但不限于:

- a) 已有的本体模型;
- b) 存储的知识单元(集);
- c) 存储的演化规则;
- d) 更新后的本体模型;
- e) 知识计算、知识获取等活动产生的新知识单元。

6.3.9.2 活动输出

知识演化活动的输出包括但不限于:

- a) 演化得出的知识单元(集);
- b) 新版本的知识图谱;
- c) 版本信息,如时间序列信息、知识源信息等。

注:为确保数据质量和避免新知识对业务的影响,对已有的知识图谱可考虑添加时间维度。

6.3.9.3 任务组成

知识演化活动的任务组成包括但不限于如下内容。

- a) 确认已有知识图谱中待变更的知识单元。
- b) 设计或匹配演化的规则、算法或模型。
- c) 更新原知识单元,如:
 - 1) 对已有知识进行变更;
 - 2) 对错误知识进行修正。
- d) 生成具有更新完成时间等信息的新版本知识图谱。

- e) 完成知识图谱版本管理,如:
 - 1) 记录知识单元变更信息和版本信息;
 - 2) 生成新知识存储路径和知识演化日志;
 - 3) 对原版本知识图谱进行冻结或淘汰。

6.3.9.4 质量一般性能

描述知识演化活动质量的一般性能可包括:

- a) 可靠性:度量是否支持多版本管控及版本的回滚、发布等;
- b) 可审查性:度量是否支持待更新知识的审查、删除等;
- c) 演化复杂度;
- d) 演化知识图谱规模;
- e) 完整性:度量更新后知识图谱和版本信息的完整程度。

6.3.10 质量保障

6.3.10.1 质量保障类型

知识图谱供应方的质量保障活动的主要类型可包括:

- a) 保障知识内容的质量,如知识的准确性、完整性、可用性、一致性、时效性等;
- b) 根据知识图谱供应方各活动的评估指标,保障知识图谱构建过程的质量;
- c) 保障知识图谱基础服务或工具的质量,如功能和性能等。

6.3.10.2 任务组成

知识图谱供应方的质量保障活动的任务组成包括但不限于如下内容。

- a) 获取质量评估需求。
- b) 完成质量检查,如:
 - 1) 实现构建过程的质量检查;
 - 2) 实现构建完成后知识图谱及其他交付成果的质量检查。
- c) 完成知识内容的纠错或更新。
- d) 完成知识图谱基础服务或工具中配置参数、功能模块、性能指标等的更新或优化。

7 知识图谱集成方

7.1 知识图谱集成方的输入

知识图谱集成方的输入包括但不限于:

- a) 知识图谱用户提供的输入:知识图谱应用需求、业务数据、辅助知识和信息系统等;
- b) 知识图谱供应方提供的输入:知识图谱基础工具、知识图谱基础服务、知识图谱技术支持等;
- c) 知识图谱生态合作伙伴提供的输入:知识图谱应用系统开发与集成所需的支撑数据、知识、技术和服务等。

7.2 知识图谱集成方的输出

知识图谱集成方的输出包括但不限于:

- a) 向知识图谱用户提供的输出:应用系统、服务支持(如咨询服务、运维服务、应用服务、系统支持)等;

- b) 向知识图谱供应方提供的输出:知识图谱构建需求、行业知识、质量指标等。

7.3 知识图谱应用系统主要构成

知识图谱应用系统是指整合知识图谱和业务实现逻辑,并具备知识图谱应用完整信息流和预设功能的系统。知识图谱应用系统主要包括以下部分。

- a) 业务应用模块,其主要功能包括:
 - 1) 提供业务数据处理、任务管理等服务;
 - 2) 提供业务功能调用及接口开放等服务。
- b) 知识图谱基础工具或基础服务[见 6.2 b) 和 c)]。
- c) 业务整合模块,其主要功能包括:
 - 1) 按照业务逻辑集成和封装业务应用模块和基于知识图谱的基础工具或服务;
 - 2) 对外提供封装后的能力调用服务。
- d) 应用交互模块,其主要功能包括:
 - 1) 提供知识呈现服务;
 - 2) 提供基于知识图谱的知识应用与交互服务。
- e) 支撑模块,其主要功能包括:
 - 1) 提供系统安全和日志审计;
 - 2) 提供用户管理和权限控制;
 - 3) 提供监控及运行支撑;
 - 4) 提供质量控制支撑;
 - 5) 提供运营及能力支撑。

7.4 知识图谱集成方的主要活动

7.4.1 活动流程

知识图谱集成方的活动流程见图 9,可包括:

- a) 需求分析活动:对知识图谱用户提出的知识图谱应用系统建设的业务需求进行分析和转换;
- b) 系统设计活动:完成知识图谱应用系统及其集成方案设计;
- c) 知识图谱集成活动:将多个知识图谱合成为一个完整和统一的知识图谱,或获取多个知识图谱的知识合成新知识;
- d) 知识图谱应用系统开发活动:完成知识图谱应用系统中算法模型与功能模块的开发,并完成模块间及与外部系统的集成;
- e) 系统维护活动:对即将或已交付的知识图谱应用系统及其集成的知识图谱和运行环境等进行配置、监控、更新和维护并保障系统的持续稳定运行;
- f) 质量保障活动:基于应用需求建立实施过程的质量监控措施,并评估实施结果的质量。

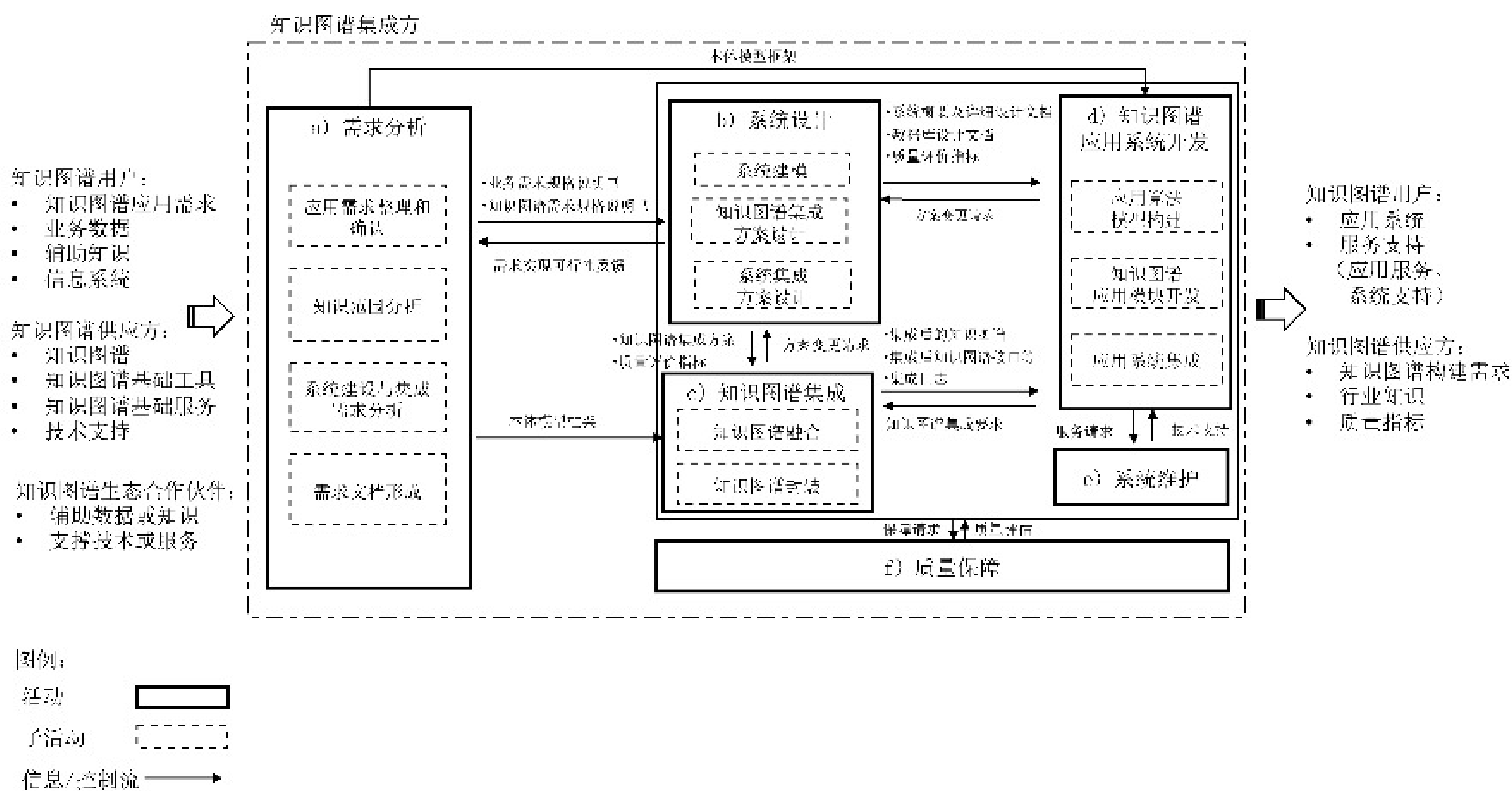


图 9 知识图谱集成方活动流程图

7.4.2 需求分析

7.4.2.1 活动输入

需求分析活动的输入包括但不限于：

- a) 知识图谱应用需求，如知识图谱应用系统建设目标、人机交互界面、知识可视化效果等；
- 注：应用需求可通过需求说明书等书面文件方式提交至知识图谱集成方进行确认。
- b) 业务数据；
- c) 辅助知识；
- d) 需求实现可行性反馈。

7.4.2.2 活动输出

需求分析活动的输出包括但不限于：

- a) 本体模型框架；
- b) 业务和系统规格说明书，可涉及业务目标、数据处理需求、业务逻辑流程图、术语定义、安全性需求、软件需求说明、功能需求、性能指标、功能要求、接口要求、服务要求、维护要求等；
- c) 知识图谱需求规格说明书，可涉及知识范围、数据需求、知识图谱应用系统验收标准、行业知识数据格式、本体模型和算法模型要求、维护要求等。

7.4.2.3 任务组成

需求分析活动中的任务组成包括但不限于如下内容。

- a) 整理和确认知识图谱用户的应用需求，如：
 - 1) 收集和整理知识图谱用户的应用需求及建设基础，如业务需求、应用场景、业务数据、专家知识等信息；
 - 2) 与知识图谱用户进行沟通和确认，形成完整的需求清单。
- b) 分析知识范围，如：

- 1) 统一建设过程中所需的术语；
- 2) 确定本体模型框架。
- c) 分析系统建设与集成需求,如:
 - 1) 定义系统集成业务需求,如业务问题、业务目标、任务明细等；
 - 2) 制定测试用例和测试计划。
- d) 形成最终完整、准确的需求文档材料,如业务和系统规格说明书、知识图谱需求规格说明书等。

7.4.2.4 质量一般性能

描述需求分析活动质量的一般性能包括但不限于：

- a) 准确性:度量需求说明书与用户需求的符合程度；
- b) 可跟踪性:度量需求分析和拆解过程的记录程度或可追溯程度；
- c) 完整性:度量需求说明书对用户需求范围的涵盖程度；
- d) 可验证性:度量需求说明书中描述的各项需求的可验证程度；
- e) 可实现性:度量在规定的时间内、经费预算和项目技术约束下开发与集成工作目标可实现程度。

7.4.3 系统设计

7.4.3.1 活动输入

系统设计活动的输入包括但不限于：

- a) 业务需求说明书；
- b) 知识图谱需求说明书；
- c) 知识图谱基础工具或服务；
- d) 知识图谱生态合作伙伴接入说明,如在系统建设过程中引入的生态合作伙伴数据、工具、服务能力说明；
- e) 方案变更请求。

7.4.3.2 活动输出

系统设计活动的输出包括但不限于如下内容。

- a) 系统概要及详细设计文档,如:
 - 1) 系统业务数据字典文档；
 - 2) 系统模块定义说明文档,如根据系统功能组成及关系定义系统的功能模块；
 - 3) 控制流图及说明文档。
- b) 数据库设计文档,如:
 - 1) 知识图谱应用系统实体关系图及说明文档；
 - 2) 知识图谱应用系统数据流图及说明文档。
- c) 知识图谱集成方案,如:
 - 1) 知识图谱集成框架:知识图谱集成目标范围内的领域概念、术语,统一知识的语义理解；
 - 2) 知识图谱集成规范:知识图谱集成过程中需要遵守的规则、约束；
 - 3) 知识图谱集成范围:知识图谱集成的边界、范围和限制。
- d) 质量评价指标。

7.4.3.3 任务组成

系统设计活动的任务组成包括但不限于如下内容。

- a) 完成系统建模,如:
 - 1) 将业务需求转化为系统需求,划分系统结构、子系统及相互间交互接口;
 - 2) 完成系统功能设计;
 - 3) 完成数据库结构设计,如数据字典和数据数据流设计、系统各模块间控制流及状态设计等。
- b) 完成知识图谱集成方案设计,如:
 - 1) 梳理或定义领域术语体系;
 - 2) 定义知识图谱集成范围;
 - 3) 设计本体模型和实例集成方案;
 - 4) 设计集成后知识单元存储方案;
 - 5) 设计知识图谱演化机制和服务接口;
 - 6) 定义知识图谱集成的辅助数据、控制参数;
 - 7) 定义知识图谱集成质量要求,编制并优化集成方案。
- c) 完成系统集成方案设计,如:
 - 1) 设计被集成系统范围、集成方式和集成计划等;
 - 2) 设计知识图谱应用系统内各子系统间、系统与外部系统间数据、功能交互流程;
 - 3) 完成接口定义;
 - 4) 制定系统集成验收要求;
 - 5) 定义系统集成评价体系。

7.4.3.4 质量一般性能

描述系统设计活动质量的一般性能包括但不限于:

- a) 完备性:度量设计内容对知识图谱集成需求、系统集成需求的覆盖程度;
- b) 可读性:度量设计内容描述清晰和无歧义的程度;
- c) 可行性:度量设计方案指导系统实现的可操作性。

7.4.4 知识图谱集成

7.4.4.1 活动输入

知识图谱集成的输入包括但不限于:

- a) 知识图谱供应方提供的两个或以上的知识图谱;
- b) 需求分析活动形成的本体模型框架;
- c) 辅助数据,如知识图谱间相同实体、关系的注释或说明等;
- d) 业务数据;
- e) 集成工具,如本体匹配工具等;
- f) 控制参数,如阈值、判断条件、版本号等。

7.4.4.2 活动输出

知识图谱集成的输出包括但不限于:

- a) 集成后的知识图谱;
- b) 集成后知识图谱的 SDK、API、可视化工具等;
- c) 集成过程数据,如有关集成过程各类详细信息等。

7.4.4.3 任务组成

知识图谱集成活动中的任务宜主要包括如下内容。

- a) 知识图谱融合：将不同知识图谱的知识单元在同一框架下进行对齐、合并和消歧，得到全局统一的知识标识和关联，如：
 - 1) 对多个知识图谱的知识单元进行归一化处理以提高后续融合效率；
 - 2) 定义知识单元或本体模型的对齐、合并、关联、消歧等规则或算法等；

注 1：面向特殊领域，可对不适于或缺失已有规则和约束的场景设计规则。

 - 3) 执行知识单元的融合流程；
 - 4) 评估知识图谱的融合效果。
- b) 知识图谱封装：将知识图谱的检索、推理等服务能力进行封装，以接口、工具或界面的形式对外输出，如：
 - 1) 确定封装形式，如：SDK、API 或可视化组件等；
 - 2) 确定封装内容并完成封装，如：知识图谱及其检索、编辑等基础能力，算法模型，计算资源调度能力等；

注 2：算法模型可包括：知识统计、知识推理、图算法等相关算法模型。

 - 3) 完成封装效果测试与优化。

7.4.4.4 质量一般性能

描述知识图谱集成活动质量的一般性能包括但不限于：

- a) 集成效率：度量集成过程中时间占用、资源占用等情况；
- b) 集成幂等：度量每次输入相同的知识图谱进行集成均可得到相同的结果；
- c) 准确度：度量实际集成效果与理想集成效果间的匹配程度；
- d) 兼容性：是否支持常规的知识图谱数据格式、集成工具；
- e) 扩展性：是否支持知识图谱全量集成、增量集成或者部分存量集成。

7.4.5 知识图谱应用系统开发

7.4.5.1 活动输入

知识图谱应用系统开发活动的输入包括但不限于：

- a) 系统设计活动形成的相关方案；
- b) 需求分析活动形成的本体模型框架；
- c) 知识图谱集成活动形成的知识图谱及集成日志；
- d) 知识图谱用户提供的业务数据和待集成信息系统；
- e) 知识图谱供应方提供的技术支持。

7.4.5.2 活动输出

知识图谱应用系统开发活动的输出包括但不限于：

- a) 知识图谱集成方案、系统集成方案变更请求；
- b) 知识图谱的应用算法模型；
- c) 集成后的知识图谱应用系统；
- d) 开发过程文档，如集成报告、问题/异常/故障报告、变更报告、验收报告等。

7.4.5.3 任务组成

知识图谱应用系统开发活动的任务组成包括但不限于如下内容。

- a) 基于应用需求开发应用算法模型,如:
 - 1) 完成应用算法模型类型和框架选型;
 - 2) 完成应用算法模型调优和迭代;
 - 3) 完成应用算法模型测试和封装。
- b) 设计并开发基于知识图谱的应用功能模块,如:

注:根据应用范围的差异,可分为知识图谱通用应用模块开发和知识图谱行业应用模块开发。其中,知识图谱通用应用模块指使用知识图谱和相关技术所生成的适用于各业务领域的应用,如知识问答、语义搜索、智能推荐等;知识图谱行业应用模块是指面向特定行业领域需求的知识图谱应用模块。

- 1) 设计应用模块业务处理逻辑、输入和输出要素等;
 - 2) 根据应用系统设计,对应用模块的输入、输出标准进行定义;
 - 3) 对知识图谱应用模块运行所需的软硬件环境、配置工具等进行预置准备;
 - 4) 完成应用模块编码、测试和封装。
- c) 完成知识图谱应用系统内部模块及外部系统的集成,如:
 - 1) 制定集成计划并按照集成设计方案实施系统集成;
 - 2) 制定集成后的系统测试用例和测试计划,并对测试未通过的部分反馈信息至集成设计活动;
 - 3) 根据测试结果,迭代优化知识图谱集成方案和系统集成方案;
 - 4) 根据集成需求、集成方案和测试报告,组织利益相关方对拟交付的知识图谱应用系统和系统支持服务方案进行验收。

7.4.5.4 质量一般性能

描述知识图谱应用系统开发活动质量的一般性能包括但不限于如下内容。

- a) 完整性:度量形成的知识图谱应用系统在不同场景下对业务需求的覆盖程度。
- b) 安全性:度量形成的知识图谱应用系统对知识、系统功能与性能、支撑软硬件设施等的保护能力,可包含完整性、连续性、保密性、可用性、可控性、透明性、隐私等,如:
 - 1) 建立有完善的知识认证和授权机制;
 - 2) 保证系统内知识图谱的完整性,保证知识单元的存储安全和传输安全;
 - 3) 保证系统中应用模块安全,如防攻击和篡改等。
- c) 可靠性:度量形成的知识图谱应用系统在指定条件下完成所需功能性和非功能性要求的能力。
- d) 响应性:度量形成的知识图谱应用系统在收到请求后返回的结果及返回结果的过程所表现出来的能力,可包括时间及资源的消耗、反馈结果的质量等。
- e) 可移植性:度量形成的知识图谱应用系统在不同软硬件环境间移植的能力。
- f) 易用性:度量形成的知识图谱应用系统在指定条件下使用时,知识图谱被理解、学习、使用和吸引用户的能力。
- g) 依从性:度量形成的知识图谱应用系统及其中知识图谱对特定领域法规、标准、伦理规则等的依存度。

7.4.6 系统维护

系统维护活动的任务组成包括但不限于如下内容。

- a) 完成知识图谱的维护。
- b) 完成应用算法模型的维护,如:
 - 1) 对已部署的算法模型使用进行状态、性能等检测,可包括知识图谱算法模型等;
 - 2) 根据接入数据、应用场景等变化对知识图谱应用系统中算法模型进行更新、调优、维护和升级。
- c) 完成分布式部署能力的扩展与维护,如:
 - 1) 对知识图谱应用系统的分布式部署环境进行状态检测;
 - 2) 根据知识图谱应用系统应用需求的扩充或变化,维护或提升分布式部署的能力。
- d) 完成知识图谱应用系统的日常维护,如:
 - 1) 针对检索访问量过大等导致系统宕机风险的情况制定维护策略;
 - 2) 定期对应用系统的运行状态和关键性能指标等进行监测,并对故障进行预警或通告。
- e) 完成知识图谱应用系统的升级,如:
 - 1) 根据日常维护情况对应用系统中存在的漏洞开发并部署系统补丁;
 - 2) 根据新出现的应用需求对知识图谱应用系统所需的数据处理能力进行扩展与维护,如数据接入、知识存储、知识传输、知识计算等;
 - 3) 根据新出现的应用需求对知识图谱应用系统的知识应用功能进行扩展与维护,如知识可视化、知识调用、拟开发的通用应用或细分领域应用等。

7.4.7 质量保障

质量保障活动的任务组成包括但不限于如下内容。

- a) 完成知识图谱应用系统集成过程的质量保障,如:
 - 1) 根据知识图谱集成方各活动的评估指标评估知识图谱应用系统开发和集成过程的质量;
 - 2) 保障知识图谱应用系统的集成耦合性、开发过程的可维护性和安全性。
- b) 完成知识图谱应用系统的功能和性能质量保障,如:
 - 1) 保障知识图谱应用系统功能的完备性、可靠性等;
 - 2) 保障知识图谱应用系统的响应时间等关键性能指标;
 - 3) 针对系统中可能存在的质量风险点制定相应的保障措施。
- c) 完成知识图谱应用系统中知识的质量保障,如:
 - 1) 保障多知识图谱集成后知识图谱的质量,如知识一致性、准确性、完整性、可用性、时效性等;
 - 2) 根据用户应用需求相关知识或其拥有知识的修正、维护或扩展对知识图谱应用系统中的知识图谱进行更新。
- d) 知识图谱应用系统服务质量保障,如:
 - 1) 实现集成后知识图谱的版本管理并提供必要的接口、工具包等服务;
 - 2) 提供知识图谱应用系统服务说明相关文档,如规范文档、过程文档、术语体系文档等。

8 知识图谱用户

8.1 知识使用者

8.1.1 知识使用者的输入

知识使用者的输入包括但不限于:知识图谱应用系统的知识单元及其他知识应用服务等。

8.1.2 知识使用者的输出

知识使用者的输出包括但不限于：应用需求，如业务应用需求、服务需求等。

8.1.3 主要活动

知识使用者的活动可包括：

- a) 调取并使用所需的知识单元；
- b) 针对特定行业，知识图谱的使用者受到的约束但不限于使用权限、知识可视化形式、知识查询精度、知识演化速度等。

注：以医学知识图谱应用为例，知识图谱为用户提供的知识受限于医疗健康领域，而不能提供诸如投资理财等金融知识。

8.2 知识维护者

8.2.1 知识维护者的输入

知识维护者的输入包括但不限于：知识图谱应用系统的知识单元及质量控制模块等。

8.2.2 知识维护的输出

知识维护者的输出包括但不限于：新增、更新、备份或恢复后的知识单元，知识维护需求等。

8.2.3 主要活动

知识维护者的活动可包括：

- a) 知识运维：对错误知识进行排除，并完成知识图谱中知识单元的新增、更新、删除等操作；
- b) 知识质量管理：对给定知识条目的内容质量进行审查，避免不合格的知识进入知识图谱，并且对质量管控严格的行业可单独设置知识质量管理者子角色；构建基于知识图谱的知识库，通过人机接口进行知识的输入和维护；
- c) 知识运营和知识推广等；
- d) 针对特定行业，知识维护者开展知识质量管理活动时宜具备相应的专业资质，如医疗、健康、金融、法律、科技、电力等。

8.3 知识提供者

8.3.1 知识提供者的输入

知识提供者的输入包括但不限于：

- a) 知识图谱基础工具；
- b) 知识图谱应用系统。

8.3.2 知识提供者的输出

知识提供者的输出包括但不限于：

- a) 知识单元，可包括原始知识、经过初步抽象和一定组织形式的常识或特定领域知识等；
- b) 知识录入需求。

注 1：知识提供者可输出知识单元至知识图谱应用系统，也可直接输出知识单元至知识图谱。

8.3.3 主要活动

知识提供者的活动可包括：

- a) 从实际知识中提取概念,得到基本概念、子问题和信息流的特征;
- b) 从业务数据中抽取符合定义的各知识要素;
- c) 参与知识图谱构建所需知识表示形式的定义或选择;
- d) 参与知识图谱构建所需本体模型的构建;
- e) 针对特定领域,知识提供者应具备与知识图谱应用系统服务领域相符的知识储备;

注 2: 对于医疗等特定行业,知识提供者还需具备相应的专业资质,并以诊断标准、专家共识等规范为基础。

- f) 知识提供者提供的知识宜符合可接受性、相关性、逻辑性、专业性等原则,其中:
 - 1) 可接受性:知识提供者提供的知识宜在认知水平下被多数人认为是正确的;
 - 2) 相关性:知识提供者提供的知识宜保证和约定的输入范围的相关性;
 - 3) 逻辑性:知识提供者提供的知识间关系逻辑宜符合客观事实;
 - 4) 专业性:知识提供者所提供的知识在其特定领域宜具备必要的专业性。

9 知识图谱生态合作伙伴

9.1 知识图谱生态合作伙伴的输入

知识图谱生态系统合作伙伴的输入包括但不限于:知识图谱供应方、集成方和用户在活动执行过程中的支撑需求等。

9.2 知识图谱生态合作伙伴的输出

知识图谱生态合作伙伴的输出包括但不限于:

- a) 技术产品:用于支撑知识图谱构建和集成的数字基础设施产品、技术工具和数据资源等,如图数据库、基础数据库或知识库和数据治理工具等;
- b) 技术服务:可包含知识图谱构建相关的支持技术(如文本识别与处理、图像处理、模式识别等)、数据治理技术服务、安全评估技术服务及评估认证服务等;
- c) 解决方案:用以指导特定行业知识图谱产品和服务开发、部署、运营和维护等相关解决方案;
- d) 咨询评估报告:根据提供的咨询服务形成的技术或商业评估报告。

9.3 主要活动

知识图谱生态合作伙伴的活动可包括:基础设施供应、数据供应、数据治理服务、咨询服务、安全服务、评估认证服务等。对应上述活动,知识图谱生态合作伙伴可划分为基础设施供应方、数据供应方、数据治理服务方、咨询服务供应方、安全服务供应方、评估认证服务供应方等子角色,见附录 A 给出的子角色说明。

附 录 A

(资料性)

知识图谱生态合作伙伴子角色说明

知识图谱生态合作伙伴的子角色包括但不限于：

- a) 基础设施供应方：为知识图谱构建及应用提供所需的软硬件基础设施，如硬件设备、云服务资源、大数据存储计算平台以及中间工具插件等；
- b) 数据提供方：将数据引入到知识图谱应用系统中并供其发现、访问、转换和展示；
- c) 数据治理服务方：负责评估、指导和监控知识图谱相关数据及其应用过程，以确保最大化数据资产的价值的同时运营合规、风险可控；
- d) 咨询服务供应方：提供知识图谱技术、应用与成熟度评价所需要的专家知识，确定知识图谱的适用性、应用范围、应用程度以及与其他信息技术的关联关系，从而实现对知识图谱技术的普及与推广；
- e) 知识图谱安全服务方：可提供知识图谱安全相关保障措施制定及部署，可包括标签数据/原始数据、获取知识、知识图谱构建的安全与隐私保护以及知识图谱应用系统的操作安全、外部接口与访问管理、系统性能的可信度管理等；
- f) 知识图谱评估认证服务方：可对知识图谱相关产品、系统或服务的质量提供检测、评估和认证服务，并可对知识图谱相关项目实施过程的质量提供评估和监理服务。

参 考 文 献

- [1] GB/T 5271.28—2001 信息技术 词汇 第28部分:人工智能 基本概念与专家系统
 - [2] GB/T 23703.2—2010 知识管理 第2部分:术语
 - [3] GB/T 40216—2021 智能仪器仪表的数据描述 属性数据库通用要求
 - [4] ISO 13374-2:2007 Condition monitoring and diagnostics of machines—Data processing, communication and presentation—Part 2: Data processing
 - [5] ISO/IEC 21838-1:2021 Information technology—Top-level ontologies (TLO)—Part 1: Requirements
-