

知识图谱构建管理系统比较研究与优化构想

马玮璐¹, 鲜国建^{1,2}, 赵瑞雪^{1,3}, 李娇^{1,3}, 黄永文^{1,3}, 孙坦^{2,4*}

(1.中国农业科学院农业信息研究所, 北京 100081; 2.农业农村部农业大数据重点实验室, 北京 100081;

3.国家新闻出版署农业融合出版知识挖掘与知识服务重点实验室, 北京 100081; 4.中国农业科学院, 北京 100081)

摘要: [目的/意义]知识图谱作为人工智能时代的重要基石, 为知识提供了一种新型组织与表示形式, 而如何高效构建并合理地管理知识图谱成为当前图谱研究人员的迫切需求。研究聚焦于已有的知识图谱构建管理系统, 以期对多款已有系统做全面深入的比较后, 总结出当前知识图谱构建管理系统的建设新思路, 并为更加通用、实用、好用的知识图谱构建管理系统研发提供参考。[方法/过程]目前大量学者针对知识图谱核心构建流程提出了先进的算法与技术, 众多知识图谱相关机构也研发了多种类型的知识图谱构建管理系统, 文中选择具有代表性的6款国内外主流知识图谱构建管理系统进行调研, 分析各系统在业务流程中的系统特色, 在系统的构建流程支持、技术选型及可用性等方面进行总结对比, 并围绕当前用户对于知识图谱构建管理系统的最新需求总结已有系统存在的局限。[结果/结论]在深入对比分析的基础上, 文中研究了一体化知识图谱协同构建管理系统的建设模式, 总结并提出分布式协同构建、多图谱并行管理、多路径知识抽取、多类图存储引擎以及跨媒体与多模态知识图谱等知识图谱构建管理系统建设的优化构想。

关键词: 知识图谱; 协同并行; 多路径抽取; 多图引擎; 管理系统

中图分类号: G251

文献标识码: A

文章编号: 1002-1248 (2023) 04-0019-13

引用本文: 马玮璐, 鲜国建, 赵瑞雪, 等. 知识图谱构建管理系统比较研究与优化构想[J]. 农业图书情报学报, 2023, 35(4): 19-31.

1 引言

知识图谱作为中国“人工智能2030”的重点发展方向之一, 拥有强大的信息检索能力、语义分析能力

和知识表示及推理能力等, 是实现认知智能的主要途径^[1], 也是语义网络发展的重要技术手段^[2]。知识图谱的构建与管理已成为当前研究和实践热点之一。

目前互联网数据量与科技信息量迅速增长, 数据存在规模大、类型多和分布领域广等特点, 这对有检

收稿日期: 2023-03-11

基金项目: 国家社会科学基金项目“融合多种知识组织体系的认知搜索模式研究”(20BTQ014); 国家科技图书文献中心(NSTL)项目“下一代开放知识服务平台关键技术优化集成与系统研发”(2023XM42)

作者简介: 马玮璐(1999-), 硕士研究生, 研究方向为知识图谱。鲜国建(1982-), 博士, 研究员, 研究方向为大数据融汇治理与知识图谱。赵瑞雪(1968-), 博士, 研究员, 研究方向为农业信息管理系统。李娇(1989-), 博士, 馆员, 研究方向为知识组织与知识图谱。黄永文(1975-), 博士, 研究员, 研究方向为知识组织与知识服务

*通信作者: 孙坦(1970-), 博士, 研究馆员(二级), 研究方向为数字信息描述与组织。E-mail: suntan@caas.cn

索、获取和利用信息等方面需求的用户造成了较大的困扰^[1],而知识图谱的出现则极大地改善了这一现象。知识图谱可以看作是一个结构化的知识库,旨在通过符号的形式来描述现实世界中事物的概念及其之间的关系,自2012年Google提出知识图谱(Knowledge Graph)^[4]以来,其在提高搜索效率,改善搜索引擎返回结果等方面发挥了重要作用。知识图谱作为大数据时代、人工智能领域的一大热点,自2009年CiteSpace^[5]传入中国以来,以可视化分析为主的科学知识图谱在图书情报领域悄然兴起^[6];同时随着关联数据及大规模开放知识库的发展,语义知识图谱在国内也获得了极大的发展^[7],现如今在医学^[8]、情报学^[9]、军事学^[10]和图书馆学^[11]等领域已有了深入的研究和应用,多数领域已构建出其专有的知识图谱,但目前由于底层数据来源不一致、构建流程不统一、构建人员间缺少协同等因素,此类语义知识图谱的构建过程繁琐费时并需要消耗大量资源对图谱进行管理。

如今国内外已有部分企业和学术机构,为便于对语义知识图谱的构建实现流程化管理并能随时根据自身需求设计构建不同的知识图谱以支撑其上层推理应用,开发了其知识图谱构建管理系统(Knowledge Graph Construction Management System, KGCMS)。然而目前大多数像图书馆等信息机构,对于知识图谱的构建仍处于“何时需要何时构建”的状态,且已有的知识图谱多面向各专业领域,缺少统一的科研知识图谱,特别是图书情报领域内现存的大量图书文献及档案等资源难以转化为更加规范的结构化知识库^[12],部分原因在于多数文献工作者以知识图谱的分析应用为主,缺少对相关图谱工具的建设研究,故如何使用一套适应性强、统一且通用的图谱系统来高效构建、管理并存储行业所聚焦的语义知识图谱成为亟待解决的重要问题。笔者依据行业标准与榜单,结合研究需求,选定了6款国内外主流的KGCMS进行系统调研,梳理了不同KGCMS在支持知识图谱构建业务流程方面的系统特色并在技术选型与可用性等方面对各系统作对比分析,总结各系统的优劣。最后针对已有KGCMS仍存在的局限,探索行业内KGCMS未来发展方向,

结合图谱相关人员现实需求,按照已发布的《知识图谱构建平台认证技术规范》^[13]相关细则,在知识图谱标准构建流程的基础之上,总结并提出了一体化知识图谱协同构建管理系统的多角度优化构想,以期完善传统的图谱系统建设方案,为行业提供构建参考。

2 主流知识图谱构建管理系统调研分析

当前,AIGC技术成为全球研究的热点,知识图谱作为支撑AIGC技术的基础也得到了快速发展^[14],在国内外已有多家机构开发了其KGCMS,可帮助机构内部快速实现知识图谱的开发构建及后续应用。本文参考《互联网周刊》公布的“2022人工智能分类排行”榜单中的“知识图谱企业排行”分支^[15],结合已有系统的受众度、系统相关文献及报道介绍,从行业内筛选出PoolParty、Hume、RelationalAI、华为云知识图谱服务平台、百度知识图谱平台和gBuilder共6款国内外主流的KGCMS,对系统的基本概况、功能流程支持、技术选型及可用性等方面进行对比分析,并结合用户当下重点关注的KGCMS建设需求对6款系统进行总结分析。

2.1 基本概况

(1) PoolParty^[16]。PoolParty是企业知识图谱实现数字化转型的平台,采用模块化的架构设计,具有文本挖掘、图谱自动化构建、图谱管理、语义AI应用等模块,其语义套件提供了必要的工具和算法,并内嵌如辞典管理、知识抽取等多项处理器,使平台可高效处理多项图谱任务。在具体的图谱构建流程中,PoolParty的一平台特色是其提供了PoolParty GraphEditor编辑器用于负责知识图谱全生命周期的管理,编辑器可提供可视化的图谱操作流程和批量、内联级的大规模图查询功能,此外PoolParty还可为每个项目定义多个PoolParty GraphEditor编辑器并授予每个编辑器不同的数据访问方式与用户权限,可有效提升知识图谱的构建效率与质量。目前PoolParty主要面向的领域有企业级知识图谱构建、语义检索、智能问答等。

(2) Hume^[17]。Hume 是 GraphAware 开发的图探索和分析工具, 具有 Schema 设计、数据映射、数据标注、图谱构建等模块。Hume 在知识图谱管理流程中的主要特点是其对图谱生命周期的各阶段允许不同用户操作, 支持不同用户间协作构建知识图谱, 并基于此可为用户提供企业级部门间的图形分析解决方案。Hume 已具备多项机器学习、知识图谱及图形可视化技术, 并依托于 Neo4j^[18]中无模式的数据模型, 可允许用户在模式层创建相同的人名、地名或物名等并为每个标签自动分配单独的 ID 值, 极大地简化了用户模式层的开发。Hume 主要面向的领域有金融、法律等企业级知识图谱构建、文本挖掘等。

(3) RelationalAI^[19]。RelationalAI 是基于关系知识图谱的下一代智能数据应用数据库系统, 其提供了一个知识图谱构建管理系统, 建设思路是将知识图谱作为一个可执行的数据和逻辑组合的模型, 并能通过声明性的程序对模型(图谱)进行扩充, 系统目前具有领域知识建模、异构数据转换集成入图、知识图谱增量更新等功能。在构建领域知识图谱模型时, RelationalAI 可提供其开发的声明性、简单易读的关系语言 Rel 来进行建模, 用户在建模时只需编写简单的 Rel 语句指定意图而无需处理过程中的各种逻辑和控制流, 建模后的图谱模型可按照既定规则直接导入各类数据从而建立起大规模的知识图谱, 简化了用户的开发编程复杂度与图谱的构建业务流。RelationalAI 目前主要服务于电信、生命科学及金融等领域的智能数据应用。

(4) 华为云知识图谱服务平台^[20]。华为云知识图谱服务平台是全流程的知识图谱构建及应用平台, 平台融合了多项图谱技术, 并提供知识图谱的具体应用场景, 可使用户快速实现知识图谱从开发到投入使用的全过程。平台内具体分为图谱项目和图谱资产库两个管理模块, 具有本体构建、信息抽取、知识映射、知识融合、知识推理、增量更新及图谱质检等功能, 图谱项目模块主要用于构建知识图谱, 用户可在其中指定该项目的本体模型、数据导入规则、抽取模型等, 图谱资产库则包含了用户个人构建的本体库和模型库, 用户可选定自己构建或系统预置的模型对图谱项目进

行构建, 构建完成的图谱数据可存放至华为云专有的华为云对象存储服务(OBS桶)^[21], 以供后续图谱应用的开发。华为云知识图谱服务平台将知识图谱构建的各流程模块集成至统一的可视化界面中, 提供大量的抽取模板且不同模块间可循环复用, 既简化了图谱构建的操作流程, 又增强了平台对于用户的友好度, 可实现零代码程式地构建图谱。目前华为云知识图谱服务平台主要面向的业务有语义搜索与推荐系统、行业场景分析、智能问答、临床决策支持及出行助手等。

(5) 百度知识图谱平台^[22]。百度知识图谱平台隶属于百度 AI 开放平台^[22], 平台以查询性能为导向且支持多源异构数据的处理, 具有知识抽取映射、多形态知识大型自动挖掘、多图管理、异构知识存储等功能, 且已在百度 AI 开放平台分别推出了图谱构建流程中的各项服务, 如知识图谱 Schema、知识理解、知识图谱构建及应用等服务, 可使用户自定义图谱的设计模型、导入本地数据、训练抽取模型、质检图谱、推理发现新知等, 能够系统地完成知识图谱的构建管理流程。百度在此基础上还推出了百度知识图谱开放平台^[23]和 EasyData 智能数据服务平台^[24], 允许用户进行数据接入并使用百度数据平台完成数据的标注、加工等操作, 完善了知识图谱构建流程上游的数据处理操作。百度知识图谱平台目前主要面向智能对话、智能搜索、智能推荐、金融医疗和法律等企业级知识图谱构建、智能媒体等领域。

(6) gBuilder^[25]。gBuilder 是知识图谱自动化构建平台, 整体采用微服务架构, 具有项目管理、Schema 设计、知识抽取、数据库和非结构化管理等模块, 平台结合机器学习、自然语言处理、图数据库等技术, 可实现数据向知识的快速转化。gBuilder 在构建知识图谱过程中的特点是其全部采用可视化设计, 可借助拖拽的方式进行 schema 设计、通过映射的方式直连关系型数据库并具备流水线构造流程可对非结构化数据进行抽取转化入图。平台同时配有自然语言问答系统 gAnswer^[26]和原生图数据库系统 gStore^[27], 可完整地实现知识图谱的应用及存储。gBuilder 目前主要面向金

融、医疗等知识图谱图谱构建以及关联分析等领域。

2.2 对比分析

根据收集的各方文献和技术资料,以及作者团队对上述 6 款 KG CMS 的使用体验,现对各 KG CMS 的图谱构建流程支持、技术选型与可用性等方面进行对比,并做总结分析。

2.2.1 对知识图谱构建管理支持对比分析

能否完整的支持知识图谱全流程构建、具备应对各类场景业务的完整图谱 workflow,是衡量 KG CMS 建设的重要标准。目前知识图谱的构建方式分为自顶向下和自底向上两种方式,常规构建流程大致为,首先采用本体构建的方式进行知识建模^[28],其次运用知识抽取技术^[29]从不同类型的数据源中获取知识,再结合知识融合技术^[30]对获取到的知识进行整合,以消除其中的矛盾和歧义,后借助知识推理技术^[31]发现原有知识中的新知识,以对知识图谱进行补充,最后通过人工和机器的方式对构建的知识图谱完成质量评估,形成合格的知识表示与特定知识库^[32]。PoolParty 具有文本分类、语料分析、实体链接、数据提取与映射、本体管理、图谱数据管理、同义词典管理和语义查询等功能,能接收并解析外部本体文件或借助系统构建原始的本体,使用户可在定义好本体模型的基础上链接合适的图谱数据,快速导入系统数据库,帮助用户完成整个知识图谱构建生命周期并将图谱快速应用于知识挖掘与服务等领域。Hume 以传统的构建流程为主支持完整的图谱 workflow,即从可视化的图谱模型设计、知识图谱的 ETL 构建到基于文本的搜索以及数据科学的应用服务等流程,其通过高可用的连接器保证可以实时读取任何数据源的数据,并为各类型数据用例编排了从数据抽取、转换到丰富的完整 workflow,且均已通过直观的 UI 组件对构建的各流程进行了表示。RelationalAI 具备完整的知识图谱构建功能, RKGMS 为其系统中数据推出了一套专属的完整性约束 (ICs),从而使各来源数据拥有统一的转化抽取、融合推理及存储共享的流程,以最大程度地降低库内图谱数据的冗余度,此外,借助其声明式语义和增量计算的方式,系统对底

层构建好的知识图谱可做到实时的更新。华为云知识图谱服务平台提供一站式知识图谱全生命周期管理服务,包括本体可视化构建、自动化图谱流水线构建,以及图谱问答、搜索、推理等图谱应用功能,其针对知识图谱的构建具有普通配置构建和智能一键构建知识图谱两种方式,并设有事先训练好的信息抽取模型,用户可根据实际数据选择合适的方案快速构建知识图谱。百度知识图谱平台用户可根据部分自身数据,小规模地训练出知识抽取和图谱构建策略模型,后向平台中导入全量数据,平台可根据模型和数据智能一体化地进行知识图谱构建,打造了面向不同结构和模态数据的知识图谱构建生态,完善了不同阶段的图谱构建流程,可以实现知识图谱从上游数据采集到下游图谱应用的全流程构建。gBuilder 内部基于 D2RQ 平台,支持结构与非结构化数据的完整抽取流程,在构建过程中可随时使用系统自带的流程检查和测试功能来检测流程的正确性与图谱构造效果,具有可视化的各模块与任务中心,有助于图谱用户实时监测图谱各流程进度,提升了构建图谱的完整性与管理效率。

除可完整地构建知识图谱外,在图谱管理流程中的分布式协同、并行管理与多路径等系统特性也被看作是 KG CMS 建设水平的重要考量标准。PoolParty 为各类数据提供专有的数据仓库 (Data Silos)、标准规范以及描述性框架,并提供了元数据管理和数据治理服务,可以保证针对不同结构数据具有完整且严谨的图谱构建管理路径,丰富了图谱数据维度。Hume 与其他知识图谱系统的主要区别是其以分布式、协作构建知识图谱为核心,可连接到广泛且分散的数据源而建立一全面真实的数据集合,各用户在 Hume 中具有不同的图谱操作权限,用户管理的每个知识图谱都可成为其所在组织内全部成员的资产,更便于企业内部部门间数据的共建共享与图谱的并行管理。RelationalAI 支持多种路径的知识抽取,可将标注的属性图数据、RDF 三元组数据和关系数据库数据集成转换为其内部规范化的关系图谱数据模型,并能通过云原生架构有效地为不同图谱构建工作提供负载隔离,在多图谱管理方面有良好的性能表现。华为云知识图谱服务平台提供

多种知识图谱抽取、融合及应用方案,其两种零编码式的智能构建图谱方案与各本体库和模型库分库管理的模式,可以更好地实现对知识图谱的多路径构建并帮助用户有效地分管系统中不同图谱项目的资源。百度知识图谱平台结合自身研发的构建与存储技术,能够很好地应对在不同路径下进行知识抽取或转换映射问题,和开放领域中多模态、跨媒体的异构知识存储问题,使平台兼容性更强,能够适应多种现实场景需求。gBuilder 对于图谱模式层具有可视化的设计功能,且在数据层针对不同结构的数据具有不同的抽取转化流程,可实现从 MySQL^[33]、Oracle^[34]、SQL Server^[35]、PostgreSQL^[36]等关系型数据库中将数据准确地映射为 RDF 三元组数据,同时提供了大量在不同非结构化数据上预训练得出的可选模型,用户可根据数据需求,挑选出最合适的模型作为构建模型,也可以使用自己的数据训练模型进行图谱构建,丰富了图谱的构建路径,同时具有分布式的计算节点,可满足各模型的并行化运行,节省了用户的系统管理开支。

2.2.2 各系统技术选型与可用性对比分析

系统的技术选型是否成熟以及可用性方面是否安全稳定,也是当前 KG CMS 建设中的重点关注内容。

在技术选型方面,用户关注的重点为 KG CMS 中是否应用了先进的机器学习和深度学习算法、是否支持当前流行的跨媒体与多模态领域、是否具备多类图存储引擎等。PoolParty 结合先进的机器学习,自然语言处理和知识图谱技术,可以链接来自不同部门、组织和外部源的数据,以帮助企业快速高效地构建庞大的知识图谱,同时在图存储方面 PoolParty 主要支持 RDF 格式数据存储,可连接 GraphDB^[37]、RDF4J^[38]等多个数据库,实现对多个图存储引擎中数据的大规模合并。Hume 具有开箱即用的自然语言处理技术,可从大量非结构化数据中抽取出新的知识,其相关的抽取技术可极易地插入到现有的服务和语言模型中,并与机器学习方法集成,为用户提供多种且灵活的访问模式,以抽取出结构、元数据和内容并转换为知识图谱中的节点和关系,同时 Hume 以属性图为主要的存储形式,底层通过 Neo4j 图数据库进行存储管理。RelationalAI

基于已有的图数据,提供了最差情况最优连接(WCOJ)算法和自动索引选择器来高效查询图谱数据,并在本地算法库也内置了主流框架和机器学习分析算法以支持海量数据处理能力,其中包括关系加速机器学习算法、图谱分析推理算法和图谱优化等算法,在数据存储方面 RelationalAI 借助配套的 RKGMS 数据库进行存储,并支持将异构数据转换为统一的关系图谱数据模型格式进行存储。华为云知识图谱服务平台融合了大量知识图谱构建、自然语言处理和大数据存储技术,并采用云服务的方式对构建完成的知识图谱进行存储,提供标准存储、低频访问存储、归档存储、深度归档存储 4 种存储模式。百度知识图谱平台研发了众多知识图谱构建技术以独立地支持其内部图谱的构建,包括融合超链分析、页面布局分析等方法的开放知识挖掘技术,基于深度学习的知识整合与补全技术和融合推理及阅读理解技术等,同时百度底层采用自研的高可用多分片式 BGraph^[39]图数据库引擎,使用查询性能导向的存储结构设计,具有高可用、高响应、可扩展和开放性等优势。gBuilder 提供了大量的在不同数据上通过深度学习等算法预训练得出的可选模型,各模型具有不同特性,用户可根据需求挑选最合适的模型作为抽取构建模型,同时利用同团队研发的 gStore 图数据库, gBuilder 可大规模存储 RDF 三元组数据并向用户导出 RDF 三元组文件,且 gStore 支持快速的 SPARQL 查询与跨复杂数据的关联分析。

在可用性方面,用户普遍更加关注 KG CMS 是否可靠实用、易于部署和二次开发,以及国内外不同平台在语种方面的适用性。PoolParty 以商用为主,面向企业,为用户提供 API 开放接口以进行开箱即用的系统集成,使用户可以在任何时间启动和运行软件,且 PoolParty 遵循信息保护管理,可帮助用户最大程度地保护其信息资产,目前 PoolParty 支持英、德、法三类语种,并在其语义套件中适配了相应语种文本数据的抽取开发。Hume 允许用户免费部署 Hume 体验版,且用户可通过本地二次开发或集成第三方工具的方式对 Hume 的各组件进行扩展,极大地提高了 Hume 各功能的灵活性,目前 Hume 已开发至 2.18 版本,支持全英

文界面,但其设计简洁,多数操作仅通过可视化拖拽的形式即可完成,降低了用户操作的复杂度。RelationalAI的RKGMS系统目前以商用为主,有较细的资源收费粒度,且基于Web网页提供服务,个人用户暂时无法本地部署并开发使用,但RelationalAI开源了其部分功能的软件工具开发包(SDK),用户可自行导入并进行本地二次开发,且目前RKGMS仅支持英文开发界面,部分操作需通过其声明性语言Rel完成。华为云知识图谱服务平台允许普通用户在申请免费体验后获得万级边的图谱存储规模和公有库中全部预置模型的使用权限,且用户可向平台申请部分功能的API接口以供后续的本地化开发使用,目前平台配备了中英文两类语种的使用界面,并在公有库中设置了针对中英文语种的不同抽取模型,用户可根据自身需求选择合适语种进行开发。百度知识图谱平台目前在其开放平台上推出了schema设计、异构数据处理等服务,用户可在注册百度智能云账号后申请体验,能够获取到在线API、SDK等开放资源,平台目前主要面向企业提供服务,可支持中英文两种语言。gBuilder目前仅向获得系统许可证服务器同意后的用户开放使用,个人用户难以获取平台使用权限或部署平台相关功能至本地使用,平台主要支持中文语种的使用,界面模块间布局清晰,操作步骤简洁。

2.3 总结分析

通过对以上6款主流的知识图谱构建管理系统进行调研和对比分析(表1),发现目前国内外不同的KGCMS在构建流程、管理模式、应用技术、支持功能、面向领域、底层图引擎及可用性等方面各有侧重。PoolParty对于多路径构建图谱技术研究较为成熟,可提供多条知识抽取路径,允许用户针对不同结构数据选择特定的构建路线,以达到最大化利用其数据的效果。Hume拥有可视化的图探索和分析界面,与用户的交互性较好,且采取异于传统图谱项目的管理方式,支持多用户在构建图谱中的协同合作与并行管理,极大地提升了图谱构建的效率。RelationalAI依托自身开发的RKGMS进行知识图谱全生命周期的构建管理操

作,集成度较高,且系统中内置了多个知识图谱构建路径与图谱相关算法,可快速实现知识图谱的一站式构建。华为云知识服务平台可允许普通用户进行注册体验并在资源库中预制了多种智能构建路径与抽取模型可供用户自行选择,简化了用户的图谱开发流程。百度知识图谱平台内部多采用自研的大数据处理技术进行平台开发,其将各项功能转化为各项服务供不同类工作人员直接调用,支持跨媒体和多模态知识图谱构建,且底层支持多源异构数据的知识存储,使平台面向领域更加广泛。gBuilder仅面向注册用户开放,系统中针对结构化和非结构化数据预置了多种抽取模型且具有不同的构建路径,采用分布式运行模式,大大提升了系统构建图谱的效率。

然而经过总结分析,发现各KGCMS在各自系统侧重点之外仍存在一定的系统局限性,具体表现在以下3点。

(1) 图谱开发过程中用户间协作度不够。如今部分KGCMS的建设仍沿用传统系统设计理念,即系统面向每个单一用户开放,但此设计模式并不利用知识图谱的全流程开发。上述6款KGCMS中华为云知识服务平台对普通用户开放的功能权限较低,并不支持普通用户实现大规模图谱的构建以及多用户的协同合作开发,对用户的友好度较低。RelationalAI对于系统中数据的存储格式具有统一要求,不利用多用户间协同式地共建共享图谱。gBuilder对于系统内不同用户的项目也采取分隔处理,难以实现多用户间高效协同构建与并行管理知识图谱。

(2) 较少KGCMS支持跨媒体与多模态领域。由于当前时代中已有大量的视频、音频及图像等不同模态数据,且此类数据也已成为知识图谱的重要数据来源,因此若KGCMS无法支持多模态数据的抽取转化,则将在一定程度上影响构建完成后图谱的质量。其中PoolParty以商用为主,具有融合企业中多种结构数据转化路径,但系统中仍缺少跨媒体与多模态类数据的构建路径。RelationalAI主要面向结构化与非结构化的文本数据,对于跨媒体与多模态类的图谱构建,系统也暂不支持。

表 1 知识图谱构建管理系统对比表

Table 1 Comparison of knowledge graph construction management systems

相关图谱系统	系统特点	相关技术	管理模式	支持功能	底层图存储支持	可用性	主要面向领域	系统局限性
PoolParty	提供必要的工具和算法;内嵌多项功能处理器;模块化架构设计;开箱即用	机器学习技术、自然语言处理技术、知识图谱相关技术	单一用户全流程化地构建管理图谱	文本挖掘、图谱自动化构建、图谱管理、语义 AI 应用	集成 GraphDB、RDF4J 等主流图数据库;提供 PoolParty GraphEditor 工具以直接编辑 RDF 数据	商用,按功能模块收费;特定功能提供 API 接口;支持英、德、法 3 类语种	企业级知识图谱构建、语义检索、智能问答	系统图存储引擎以面向 RDF 三元组数据为主,缺少属性图类数据的存储引擎;缺少跨媒体与多模态类数据的构建路径、
Hume	图探索和分析工具;协作构建知识图谱;成员内数据高效共享	机器学习技术、知识图谱相关技术、图形可视化技术	支持多用户在图谱构建中协同合作、图谱管理中并行处理	Schema 设计、数据映射、数据标注、图谱构建	依托 Neo4j 图数据库管理	可免费部署 Hume 体验版;可通过本地二次开发或集成的方式对其中组件进行扩展;支持英语语种	金融、法律等企业级知识图谱构建、文本挖掘	底层仅提供 Neo4j 图数据库进行存储,仅支持属性图数据格式的存储
RelationalAI	基于关系知识图谱的下一代智能数据应用数据库系统;将知识图谱作为一个可执行的数据和逻辑组合的模型;具备声明性开发语言 Rel	机器学习技术、知识图谱技术、Rel 开发技术	单一用户仅可获取该账户下系统中全部资源,统一进行图谱的构建管理	领域知识建模、异构数据转换集成入图、知识图谱增量更新	借助配套的 RKGCMS 数据存储,支持将异构数据统一转换为关系图谱数据模型进行存储	商用,按使用资源类型单独收费;部分功能的 SDK 可自行本地导入并二次开发	电信、生命科学及金融等领域的智能数据应用	底层图存储引擎单一,对导入系统的异构数据有较严格规范;系统不支持跨媒体与多模态图谱构建
华为云知识图谱服务平台	一站式知识图谱全生命周期管理服务;流水线式快速构建及管理知识图谱;智能一键构建与自定义构建两种构建路径	知识图谱构建技术、自然语言处理技术、大数据存储技术、云服务技术	系统中分隔图谱项目及系统资源库,图谱项目面向单一用户构建,资源库中模型允许不同用户间共享	本体构建、信息抽取、知识映射、知识融合、知识推理、增量更新、图谱质检	华为云对象存储服务(OBS),提供标准存储、低频访问存储、归档存储、深度归档存储 4 种存储模式	普通用户可拥有 30 天有效期;平台内按功能权限收费;可基于 HTTPS 请求 API 进行部分功能的本地集成;支持中英文语种	语义搜索与推荐系统、行业场景分析、智能问答、临床决策支持、出行助手	不支持普通用户实现大规模图谱的构建;多用户间难以协同合作开发知识图谱
百度知识图谱平台	以查询性能为导向;支持多源异构数据	机器学习技术、大数据技术、云计算技术、自然语言处理与理解技术	将图谱构建中的各功能划分为各项服务,供不同角色的用户可直接调用	知识抽取映射、多形态知识大型自动挖掘、多图管理、异构知识存储	百度自研高可用多分片式 BGraph 图数据库引擎	商用,普通用户注册后可获取到在线 API、SDK 等开放资源;支持中英文语种	智能对话、智能搜索、智能推荐、金融医疗与法律行业、智能媒体	平台主要面向单一用户的图谱构建业务开发,用户间协同构建较少;系统底层仅支持单一的自研图存储引擎;
gBuilder	北京大学知识图谱自动化构建平台;微服务架构	机器学习技术、自然语言处理技术、图数据库技术	系统面向单一用户开放使用,系统中分为图谱项目构建与资源库,不同用户间资源不共享	项目管理、Schema 设计、知识抽取、数据库和非结构化管理	面向 RDF 数据存储的团队自研 gStore 图数据库	用户需经由许可证服务器申请同意后后方可使用;支持中文语种	领域图谱构建、关联分析	仅支持 RDF 三元组形式的图谱数据存储,图谱数据存储形式单一;用户可选择的图存储引擎较少

(3) 多数 KG CMS 的底层图存储引擎较单一。目前对于知识图谱的存储主要有属性图和 RDF 三元组两种存储形式, 分别对应了不同主流的图数据库, 因此当 KG CMS 底层的图存储引擎较单一时, 用户难以根据自身数据选择读写速率快且存储规模合适的图引擎。其中 Hume 底层依托 Neo4j 图数据库进行管理, 各来源数据在 Hume 中统一转化为 Neo4j 中属性图形式进行存储, 底层图存储引擎选择较少, 图谱数据的存储格式较单一。百度知识图谱平台配备了自研的 BGraph 图存储引擎, 多源异构数据在平台中需转化为 BGraph 规范的存储格式, 用户无法选择其他类型的图存储引擎。gBuilder 底层使用 gStore 进行图谱存储, 仅支持 RDF 三元组形式的图谱数据存储, 缺少图数据库和关系数据库等类型的图存储引擎。

3 一体化知识图谱构建管理系统优化建设构想

目前国内外对于 KG CMS 已有了较深入研究, 各机构推出的 KG CMS 在构建流程、应用技术和理论方法等方面趋于成熟, 正逐步打造出适合不同领域构建知识图谱所需的系统。然而用户针对系统功能模块、技术选型及应用场景等方面还有诸多需求, 根据 CESI 颁布的《知识图谱构建平台认证技术规范》, 成熟的 KG CMS 应拥有完整的知识图谱构建流程, 保证用户可以从零到一全流程式、一体化地构建知识图谱且系统的不同功能模块划分清晰, 适配灵活^[13], 对于具体的流程处理如知识抽取、知识融合、知识存储等阶段, 系统应能满足具体数据标准的测试要求。根据本文对上述 6 款 KG CMS 分析, 并结合已有学者对众多 KG CMS 在系统架构、平台功能等方面的分析^[40], 发现目前呈现出的整体趋势为多数企业推出商用的 KG CMS, 虽部分系统内均预置了多条图谱构建路径, 但其面向普通用户仅开放少数系统预定义的构建路线, 商用系统开放度较低, 在图谱的存储上用户的访问权限也较少, 图存储引擎选择较单一; 部分科研机构的 KG CMS 则主要面向特定领域开发、难以支持跨媒体与多模态领域

的图谱构建, 且较少有系统具备多用户式协同构建的开发策略; 由于多数系统主要面向图谱构建用户开发, 缺少对于如数据标注人员、图谱质检员等用户的考虑, 故无法实现 KG CMS 中多图谱的并行管理。因此, 现有的 KG CMS 在实现完整的知识图谱构建流程基础上, 很难同时做到多人协同分布式地管理图谱、多路径式地知识抽取、多图引擎存储、面向跨媒体多模态领域的图谱构建以及多角度、宽领域式地应用知识图谱。针对行业内用户在认知智能时代对于 KG CMS 的新需求及对 6 款 KG CMS 的深入对比分析, 本文从以下 5 个角度提出了更为通用、适用的一体化知识图谱构建管理系统的优化建设方案, 如图 1 所示为在新需求下的 KG CMS 的系统模型图。

3.1 分布式协同构建

知识图谱往往需要海量的多源异构数据与复杂的流程进行构建, 因此仅凭个人用户难以实现高效且准确的构建大规模知识图谱, 故更现代化的 KG CMS 应能做到多人协同式地分阶段图谱构建^[41]。在团队中可根据不同的功能模块交由不同职责的人员处理, 如图谱管理人员负责知识图谱构建的流程细节、数据标注人员负责对各类型输入数据的标签编辑、系统管理员负责日常系统的运维与服务等, 实现可分布式地构建同一知识图谱。系统的各模块也应做到可分可合, 后续用户可基于其他用户事先定义的图谱模型或抽取模型直接进行复用, 也可根据权限内其他用户的标注数据直接进行现有模型的训练, 可促进团队内不同图谱项目之间的信息共享。分布式地多人构建图谱使专门的图谱用户可更加聚焦于知识图谱构建的核心流程, 大大提升了系统构建图谱时的整体效率, 也使不同操作人员可以专精于特定的图谱构建模块, 化简繁琐的图谱构建流程, 有助于提高构建图谱的准确度。

3.2 多图谱并行管理

针对不同领域的知识图谱, 其具有不同类型的图谱模型、不同来源的图谱数据以及不同步骤的构建流程及数据管理方式^[42]等, 因此需要分设不同的图谱项

同模态数据, 目前已有大量机构将此类跨媒体、多模态数据作为知识图谱的重要数据来源。因此, 完整的 KGCMS 应具备多模态数据前期的特征对齐、联合表征等预处理环节, 使不同模态数据具备完整的特征标签, 将杂乱的多模态数据过滤成为有效且规范的图谱输入数据; 并为用户提供针对不同模态数据的多模态融合策略^[44]及知识图谱构建方法, 具备不同模态信息的机器学习模型、多种跨媒体数据的融合算法等, 且具有跨媒体式的图谱模型表示机制, 可使用户清晰地提取文本、图像、视频和音频等模态数据中的语义信息。对于处理完成的不同模态数据, KGCMS 具备不同模态数据的存储引擎, 保证大规模数据的高速读写, 在系统建立起庞大的文本、语音、图像及视频等知识库, 并为此类型数据搭建好底层存储架构与策略, 以便后续同类型数据的快速读写以及更好地支撑知识图谱后续相关的跨媒体应用。

3.5 多类图存储引擎

KGCMS 的图存储引擎管理是支撑知识图谱上层构建管理的基础, 已有学者针对知识图谱的数据库管理进行了研究设计^[45], 目前常见的知识图谱存储方式有 RDF 三元组、图数据库和关系型数据库的形式, 其中代表性的有 Virtuoso^[46]、GraphDB、Neo4j、Nebula-Graph^[47]以及 MySQL、SqlServer 等关系型数据库, 不同类型的数据库具有不同的数据存储策略、不同规模的数据存储空间、不同形式的数据组织架构和查询语言以及不同规格的数据读写标准等。由于用户图谱数据规模、类型及后续图谱服务内容的不同, 因此 KGCMS 应为用户提供多类图存储引擎, 用户可根据图谱数据来源、构建的图谱类型及规模选取合适的图引擎进行存储。同时在 KGCMS 中预置不同图引擎间数据交互转化的处理操作, 便于用户对构建完成的知识图谱在不同引擎间快速地迁移转化。在 KGCMS 中部署多类图存储引擎, 实现多种格式的数据在 KGCMS 中动态高效地分布式存储, 对系统底层的存储资源分配利用更加合理, 也更有利于支持后续多种图谱场景的应用开发。

4 结 语

本文系统地研究了国内外 6 款主流的知识图谱构建管理系统, 调研了各系统的主要特点, 在各系统对知识图谱构建及管理流程的功能支持、系统技术选型与可用性等方面展开对比分析, 梳理了各 KGCMS 的优劣, 总结出已有的 KGCMS 在当前行业中的整体趋势与现有局限。最后本文基于已有的知识图谱构建管理系统建设标准与已有学者对行业中主流 KGCMS 的深入研究, 分析图谱用户在当下日益增长的各项现实需求, 结合对已有 6 款 KGCMS 的对比分析, 并考虑目前认知智能时代的数据特点, 探讨了新型知识图谱构建管理系统的未来建设模式, 提出了支持分布式协同构建、多图谱并行管理、多路径知识抽取、跨媒体与多模态、多类图存储引擎的一站式、全流程知识图谱构建管理系统优化构想, 以期为行业研发更加一流的 KGCMS 提供建设新思路与方案参考。

参考文献:

- [1] 王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. 计算机研究与发展, 2022, 59(9): 1947-1965.
WANG M, WANG H F, LI B H, et al. Survey on key technologies of new generation knowledge graph [J]. Journal of computer research and development, 2022, 59(9): 1947-1965.
- [2] BIZER C, HEATH T, BERNERS-LEE T, et al. Linked data: The story so far[M]. USA: IGI global, 2011.
- [3] 王颖. 科技文献内容语义描述模型研究[J]. 农业图书情报学报, 2020, 32(8): 12-24.
WANG Y. Semantic Models for the Content of Scientific Literature[J]. Journal of library and information science in agriculture, 2020, 32(8): 12-24.
- [4] Official google blog: Introducing the knowledge graph: things, not strings [EB/OL]. [2023-06-08]. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.
- [5] CiteSpace [EB/OL]. [2023-05-30]. <https://sourceforge.net/projects/citespace/>.

- [6] 王东浩. 基于科学知识图谱的图书情报学科可视化比较研究——评《中外图书情报学科知识图谱比较研究》[J]. 中国科技论文, 2022, 17(10): 1180.
- WANG D H. Comparative study on visualization of library and information science based on scientific knowledge map – Comment on comparative study of knowledge map of library and information science between China and foreign countries[J]. China sciencepaper, 2022, 17(10): 1180.
- [7] 唐亮, 罗轩, 王颖. 两类知识图谱差异辨析及其在科技出版中的应用[J]. 出版参考, 2019(1): 66–68.
- TANG L, LUO X, WANG Y. Differentiation and analysis of the differences between two types of knowledge maps and their application in sci-tech publishing[J]. Publishing reference, 2019(1): 66–68.
- [8] 杨云飞, 穗志方. 面向医学知识图谱的可视化方法设计与实现[J]. 中文信息学报, 2022, 36(2): 40–48.
- YANG Y F, SUI Z F. Design and implementation of visualization for medical knowledge graph[J]. Journal of Chinese information processing, 2022, 36(2): 40–48.
- [9] 张海涛, 栾宇, 周红磊, 等. 总体国家安全观下重大突发事件的智能决策情报体系研究[J]. 情报学报, 2022, 41(11): 1174–1187.
- ZHANG H T, LUAN Y, ZHOU H L, et al. Intelligent decision-making information system for major emergencies: A holistic approach to national security[J]. Journal of the China society for scientific and technical information, 2022, 41(11): 1174–1187.
- [10] 马玉凤, 向南, 豆亚杰, 等. 军事系统工程中的知识图谱应用及研究[J]. 系统工程与电子技术, 2022, 44(1): 146–153.
- MA Y F, XIANG N, DOU Y J, et al. Application and research of knowledge graph in military system engineering[J]. Systems engineering and electronics, 2022, 44(1): 146–153.
- [11] 于升峰. 面向科技智库的知识图谱系统构建[J]. 智库理论与实践, 2021, 6(1): 56–64.
- YU S F. Construction of mapping knowledge system for science and technology think tanks[J]. Think tank: Theory & practice, 2021, 6(1): 56–64.
- [12] 唐玫, 赵婉忻, 李晶, 等. 面向科学数据管理的机构知识库建设与启示[J]. 数字图书馆论坛, 2022(10): 65–72.
- TANG M, ZHAO W X, LI J, et al. Development situation and enlightenment of the institutional repositories for research data management[J]. Digital library forum, 2022(10): 65–72.
- [13] 021–2020 C T. 知识图谱构建平台认证技术规范[S]. 北京: 中国电子技术标准化研究院赛西实验室, 2020.
- 021–2020 C T. Technical specification for certification of knowledge graph construction platforms[S]. Beijing: Chinese Electronic Standardization Institute, CESI Laboratory, 2020.
- [14] BALLARDINI R M, HE K, ROOS T. AI-generated content: Authorship and inventorship in the age of artificial intelligence[M]// Online distribution of content in the EU. England: Edward elgar publishing LTD, 2019.
- [15] 2022 人工智能分类排行[EB/OL]. [2023–03–14]. <https://baijiahao.baidu.com/s?id=1739952164626665131&wfr=spider&for=pc>.
- 2022 AI classification ranking[EB/OL]. [2023–03–14]. <https://baijiahao.baidu.com/s?id=1739952164626665131&wfr=spider&for=pc>.
- [16] PoolParty[EB/OL]. [2023–03–14]. <https://www.poolparty.biz/>.
- [17] Hume – mission-critical graph analytics [EB/OL]. [2023–03–14]. <https://graphaware.com/products/hume/>.
- [18] Neo4j[EB/OL]. [2023–05–30]. <https://neo4j.com/>.
- [19] RelationalAI: Build data apps with intelligence[EB/OL]. [2023–03–14]. <https://relational.ai/>.
- [20] 华为云知识图谱服务[EB/OL]. [2023–03–14]. <https://support.huaweicloud.com/kg/index.html>.
- Huawei cloud knowledge graph service [EB/OL]. [2023–03–14]. <https://support.huaweicloud.com/kg/index.html>.
- [21] 对象存储服务 OBS[EB/OL]. [2023–05–30]. <https://www.huaweicloud.com/product/obs.html>.
- Object storage service (OBS) [EB/OL]. [2023–05–30]. <https://www.huaweicloud.com/product/obs.html>.
- [22] 百度 AI 开放平台[EB/OL]. [2023–03–14]. <https://ai.baidu.com/solution/kgaa>.
- Baidu AI open platform[EB/OL]. [2023–03–14]. <https://ai.baidu.com/solution/kgaa>.
- [23] 百度知识图谱开放平台[EB/OL]. [2023–05–30]. <https://kgopen.baidu.com/index>.
- Baidu knowledge graph open platform[EB/OL]. [2023–05–30]. <https://kgopen.baidu.com/index>.

- [24] EasyData 智能数据服务平台[EB/OL]. [2023-05-30]. <https://ai.baidu.com/easydata/>.
EasyData intelligent data service platform[EB/OL]. [2023-05-30]. <https://ai.baidu.com/easydata/>.
- [25] 北京大学知识图谱自动化构建平台 gBuilder[EB/OL]. [2023-03-14]. <http://openkg.cn/tool/gbuilder>.
gBuilder: Beijing university knowledge graph automation construction platform[EB/OL]. [2023-03-14]. <http://openkg.cn/tool/gbuilder>.
- [26] 北京大学知识库问答系统 gAnswer[EB/OL]. [2023-05-30]. <https://answer.gstore.cn/pc/index.html>.
gAnswer: Peking university knowledge base question-answering system[EB/OL]. [2023-05-30]. <https://answer.gstore.cn/pc/index.html>.
- [27] 北京大学原生图数据库系统 gStore[EB/OL]. [2023-05-30]. [https://www.gstore.cn/pcsite/index.html/#/](https://www.gstore.cn/pcsite/index.html#/).
gStore: Peking university native graph database system [EB/OL]. [2023-05-30]. [https://www.gstore.cn/pcsite/index.html/#/](https://www.gstore.cn/pcsite/index.html#/).
- [28] KHADIR A C, ALIANE H, GUESSOUM A. Ontology learning: Grand tour and challenges[J]. Computer science review, 2021, 39: 100339.
- [29] 邓依依, 鄢昌兴, 魏永丰, 等. 基于深度学习的命名实体识别综述[J]. 中文信息学报, 2021, 35(9): 30-45.
DENG Y Y, WU C X, WEI Y F, et al. A survey on named entity recognition based on deep learning[J]. Journal of Chinese information processing, 2021, 35(9): 30-45.
- [30] SMIRNOV A, LEVASHOVA T. Knowledge fusion patterns: A survey[J]. Information fusion, 2019, 52: 31-40.
- [31] CHEN X J, JIA S B, XIANG Y. A review: Knowledge reasoning over knowledge graph[J]. Expert systems with applications, 2020, 141: 112948.
- [32] HOGAN A, BLOMQUIST E, COCHEZ M, et al. Knowledge graphs[J]. ACM computing surveys, 2021, 54(4): 71.
- [33] MySQL[EB/OL]. [2023-05-30]. <https://www.mysql.com/>.
- [34] Oracle[EB/OL]. [2023-05-30]. <https://www.oracle.com/>.
- [35] Microsoft SQL server[EB/OL]. [2023-05-30]. <https://www.microsoft.com/zh-cn/sql-server/sql-server-downloads>.
- [36] PostgreSQL[EB/OL]. [2023-05-30]. <https://www.postgresql.org/>.
- [37] GraphDB[EB/OL]. [2023-05-30]. <http://www.graphdb.net/>.
- [38] RDF4J[EB/OL]. [2023-05-30]. <https://rdf4j.org/>.
- [39] BGraph[EB/OL]. [2023-05-30]. <https://ai.baidu.com/tech/kg/bgraph>.
- [40] 王传庆, 李阳阳, 费超群, 等. 知识图谱平台综述[J]. 计算机应用研究, 2022, 39(11): 3201-3210.
WANG C Q, LI Y Y, FEI C Q, et al. Survey of knowledge graph platform[J]. Application research of computers, 2022, 39(11): 3201-3210.
- [41] LI M L, NI Z H, TIAN L, et al. Research on hierarchical knowledge graphs of data, information, and knowledge based on multiple data sources[J]. Applied sciences, 2023, 13(8): 4783.
- [42] 王鑫, 邹磊, 王朝坤, 等. 知识图谱数据管理研究综述[J]. 软件学报, 2019, 30(7): 2139-2174.
WANG X, ZOU L, WANG C K, et al. Research on knowledge graph data management: A survey[J]. Journal of software, 2019, 30(7): 2139-2174.
- [43] CHEN Z K, ZHAO Y. The technology of military knowledge graph construction based on multiple open data sources[C]// 2020 5th international conference on mechanical, control and computer engineering (ICMCCE). Piscataway, New Jersey: IEEE, 2021: 1993-1997.
- [44] GAO J, LI P, CHEN Z K, et al. A survey on deep learning for multimodal data fusion[J]. Neural computation, 2020, 32(5): 829-864.
- [45] 刘宝珠, 王鑫, 柳鹏凯, 等. KGDB: 统一模型和语言的知识图谱数据库管理系统[J]. 软件学报, 2021, 32(3): 781-804.
LIU B Z, WANG X, LIU P K, et al. KGDB: Knowledge graph database system with unified model and query language [J]. Journal of software, 2021, 32(3): 781-804.
- [46] Virtuoso[EB/OL]. [2023-05-30]. <https://sourceforge.net/projects/virtuoso/>.
- [47] NebulaGraph[EB/OL]. [2023-05-30]. <https://www.nebula-graph.io/>.

Comparative Study and Optimization Strategies of Knowledge Graph Construction Management Systems

MA Weilu¹, XIAN Guojian^{1,2}, ZHAO Ruixue^{1,3}, LI Jiao^{1,3}, HUANG Yongwen^{1,3}, SUN Tan^{2,4*}

(1. Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081; 2. Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081; 3. Key Laboratory of Knowledge Mining and Knowledge Services in Agricultural Converging Publishing, National Press and Publication Administration, Beijing 100081; 4. Chinese Academy of Agricultural Sciences, Beijing 100081)

Abstract: [Purpose/Significance] Knowledge Graph has become a major research hotspot in the era of artificial intelligence due to its ability to provide a new means of organization and representation of knowledge. As the field continues to evolve, numerous scholars have proposed advanced algorithms and technologies for each core stage of constructing a knowledge graph, and many large domestic and foreign enterprises have also developed their independent knowledge graph management systems. However, the majority of these graph tools developed are designed for commercial use and are often too expensive and difficult to deploy locally for small and medium-sized research teams. This presents a challenge for information organizations such as research libraries with massive resources, which require a more adaptable, universal, and efficient tool to build and manage knowledge graphs. To meet this need, it is important to develop an open-source, user-friendly, and customizable knowledge graph management system that can be easily deployed by small and medium-sized research teams. [Method/Process] In summary, this article offers a thorough and informative analysis of six mainstream knowledge graph management systems, both domestically and internationally. It delves into the unique characteristics of each system within the business process and provides an in-depth comparative analysis based on several important factors, including system functionality, technology selection, open-source availability, and application domains. The article refers to the standard construction process of knowledge graphs and highlights the platform characteristics of each system during the construction process while also examining their limitations based on current data characteristics. In response to practical needs, the article focuses on multi-path, multi-engine, distributed, and collaborative construction, integrating advanced graph algorithms and considering a well-developed underlying graph storage strategy. [Results/Conclusions] As a result, the article presents an in-depth analysis of the construction model for a collaborative development and management system of an integrated knowledge graph. It not only investigates the current state of knowledge graph management systems but also proposes novel optimization ideas. These ideas include distributed collaborative construction, which allows for simultaneous contributions from multiple sources, and parallel management of multiple graphs, enabling efficient organization and retrieval. Additionally, some suggestions are put forward: developing multi-path knowledge extraction techniques to enhance the knowledge acquisition process, and using specialized multi-graph storage engines for optimized storage and retrieval. Last, the article emphasizes the importance of incorporating cross-media and multimodal knowledge into the graph for a comprehensive representation of information.

Keywords: knowledge graph; collaborative parallelism; multi-path extraction; multi-graph engine; management system