



情报杂志
Journal of Intelligence
ISSN 1002-1965, CN 61-1167/G3

《情报杂志》网络首发论文

题目：基于 LLM 与可视化的图数据库专利知识图谱构建及分析
作者：周贞云，黄昭昭，邱均平
网络首发日期：2025-07-16
引用格式：周贞云，黄昭昭，邱均平. 基于 LLM 与可视化的图数据库专利知识图谱构建及分析[J/OL]. 情报杂志.
<https://link.cnki.net/urlid/61.1167.G3.20250716.1339.010>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于 LLM 与可视化的图数据库专利 知识图谱构建及分析*

周贞云^{1,2,3} 黄昭昭^{1,2,3} 邱均平^{1,2,3}

- (1. 杭州电子科技大学图技术与数字化联合实验室 杭州 310018;
2. 杭州电子科技大学数据科学与信息计量研究院 杭州 310018;
3. 杭州电子科技大学中国科教评价研究院 杭州 310018)

摘要: [研究目的] 针对传统专利知识图谱(PKG)构建方法效率低的问题,本研究提出基于大语言模型(LLM)的专利知识图谱构建方法,旨在实现图数据库领域专利信息的智能化处理,进而实现基于知识图谱的专利信息分析。[研究方法] 以图数据库专利为样本,设计 LLM 提示词抽取知识以构建 PKG。借助 Neo4j 可视化,多维度梳理技术演化路径,运用 TextRank、PageRank 算法分析应用领域关键词与节点影响力。[研究结果/结论] 研究表明,通过 LLM 能够实现自动化构建知识图谱;图数据库技术历经基础架构优化、分布式扩展到 AI 融合三阶段发展;应用领域以知识图谱、数据处理的模型为核心,其中知识图谱节点影响力最强。
关键词: 大语言模型;图数据库;专利信息;知识图谱;知识抽取;可视化
中图分类号: TP391.1;G255.53 **文献标识码:** A

Construction and Analysis of Patent Knowledge Graph in Graph Database Field Based on LLM and Visualization

Zhou Zhenyun^{1,2} Huang Zhaozhao^{1,2} Qiu Junping^{1,2}

- (1. Graph Technology and Digitization Joint Laboratory of Hangzhou Dianzi University, Hangzhou 310018;
2. Hangzhou Dianzi University, Hangzhou 310018)

Abstract: [Research purpose] Addressing the issue of low efficiency in traditional patent knowledge graph (PKG) construction methods, this study proposes a patent knowledge graph construction method based on large language models (LLM), aiming to achieve intelligent processing of patent information in the field of graph databases, and thereby enabling knowledge graph-based patent analysis. [Research method] Using graph database patents as samples, this study designs LLM prompt word extraction knowledge to construct PKG. With the help of Neo4j visualization, it sorts out the technological evolution path from multiple dimensions, and analyzes the influence of application field keywords and nodes using TextRank and PageRank algorithms. [Research result/conclusion] The research shows that LLM can achieve automated construction of knowledge graphs; graph database technology has undergone three stages of development: infrastructure optimization, distributed expansion, and AI integration; the application fields are centered around knowledge graphs, data processing, and models, with knowledge graph nodes having the strongest influence.

Key words: large language model; graph database; patent information; knowledge graph; knowledge extraction; visualization

基金简介: 国家自然科学基金重大项目“基于大数据的科教评价信息云平台构建和智能服务研究”(编号:19ZDA348);浙江省科学技术厅技术开发项目“基于图技术的科教评价信息资源开发与管理”(编号:KYH263123010M)研究成果。

作者简介:周贞云,男,1979年生,博士研究生,副教授,研究方向:数据科学与信息计量,科技创新与科学评价;黄昭昭,女,1996年生,硕士研究生,研究方向:图数据库与知识图谱;邱均平,男,1947年生,博士生导师,资深教授,研究方向:信息计量与科教评价。

通信作者:邱均平

0 引言

随着各行业对大数据分析 & 关系挖掘需求的增加^[1],传统关系数据库在存储容量、查询效率、扩展性等方面的局限性日益凸显,此状推动了新型数据库技术的发展^[2]。其中,图数据库以节点和边表示数据关系,在处理复杂关系查询中展现显著的优势^[3-4],相关专利数据量逐年上升。这些专利蕴含丰富的技术信息,对科技情报分析和决策意义重大^[5],但如何高效组织及利用海量专利信息是一个关键问题。

知识图谱技术的进步^[6]为专利组织与分析提供了新思路,专利知识图谱(Patent Knowledge Graph, PKG)通过结构化知识呈现形式,能够揭示专利关联与技术演进路径,为预测技术趋势、开展竞争情报分析等提供有力支持^[7]。当前PKG构建方法主要依赖人工标注和基础文本分析,难以满足大规模数据处理和高级应用需求。近年来,基于Transformer架构^[8]的大语言模型(Large Language Model, LLM)^[9]具备强大的自然语言处理能力,可处理复杂文本、提取潜在知识^[10],在专利文献解析和知识提取方面极具应用潜力。将LLM与PKG结合,有望提升专利信息处理的自动化水平,实现更智能的专利知识抽取。

基于此,本研究提出一种基于LLM的PKG构建方法,应用于图数据库相关专利中,并结合PKG分析图数据库技术演化及应用态势。本研究期望为图数据库相关研究提供参考,为专利信息的智能化处理与知识挖掘提供研究视角。

1 相关研究

知识抽取是构建PKG的关键,其效果直接影响图谱质量^[11],主要包括基于规则、统计机器学习、深度学习三类^[12]。这些方法在专利知识抽取领域也得到了广泛的实践研究。基于规则的方法,如韦景竹^[13]等依据关键词和术语相关度、首尾度提取专利术语;何喜军^[14]等运用正则匹配和语义相似度进行实体识别。此类方法靠专家制定规则,通过统计信息、关键字等特征提取实体^[12],但规则设计依赖大量领域知识,处理多种实体类型或面对语言差异时难度大增,难以应对未知或超出规则范围的实体。在统计机器学习方法方面,姜春涛^[15]和王密平^[16]等人利用条件随机场分别提取专利引文和术语;陈亿群^[17]等利用支持向量机自动提取专利关键词,这类方法自动化程度高但需要大量标注数据且语义理解有限。

深度学习的出现优化了专利知识抽取的表现,Hengqin Wu^[18]等基于Transformer自动识别建筑领域专利文件中的通信实体;Haixiang Yang^[19]等结合Ro-

BERTa和CNN提取专利技术点;Geng, BT^[20]提出Bi-LSTM和CNN的混合框架用于专利要求中的关系抽取。此外,不少研究结合统计机器学习与深度学习进行改进,何玉^[21]等基于SpERT架构提出SpERT-Ag-gcn信息抽取模型;何喜军^[22]等提出基于BERT-BiLSTM-CRE的语义实体识别模型。这类方法泛化能力和灵活性强,但需大量标注数据、计算资源和时间,且参数调优复杂。

近年来,Transformer架构催生的BERT^[23]、GPT^[24]等LLM被积极应用于知识抽取任务中。李佳斌^[25]等基于Xunzi_GLM基座模型整合词典知识,裴炳森^[26]等则结合Chinese Llama2-7B模型与小样本提示在公安执法文本进行知识抽取。在专利领域,LLM也已应用于相关任务,赵建飞^[27]等基于Baichuan2-7B-Chat的LLM实现无标注的专利关键信息抽取,张玲玲和黄务兰^[7]利用ChatGPT API抽取专利摘要信息。这些方法无需大规模标注数据,且性能优于其他方法。

总体而言,基于规则和统计机器学习的知识抽取方法在特定情境下仍具价值。但深度学习大幅提升了抽取的准确性与效率,LLM更是降低对大量标注数据的依赖。将LLM用于专利知识抽取,能提高自动化程度,突破数据标注瓶颈,在PKG构建中潜力巨大。现有研究中,PKG构建与应用的探讨相对成熟^[28-30],但缺乏图数据库领域的PKG研究。鉴于此,本研究设计并实施基于LLM和提示词工程^[31]的知识抽取策略,对专利文本开展自动化实体抽取。同时,结合图数据库相关专利信息与知识图谱技术,深入剖析图数据库技术的演化进程及应用态势,期望全面揭示图数据库技术发展动态,为相关领域研究提供新的见解。

2 研究设计

专利知识图谱的构建与分析的层次可分为模式层、数据层和应用层。模式层预定义本体,指导数据层的知识组织与存储;数据层获取实体数据,依本体映射、组织后集成至PKG;应用层利用PKG开展图数据库领域专利分析。本文聚焦图数据库领域的PKG构建,采用自顶向下的构建策略,以图数据库相关专利摘要为构建对象,整体框架如图1。

在模式层设计上,结合图数据库相关专利的知识背景和已有研究对专利摘要要素的归纳,定义实体、属性和关系,构建PKG本体。在数据层构建时,从专利数据库采集数据并进行预处理,然后利用LLM抽取知识,生成半结构化数据(JSON格式)并转换为结构化数据,通过知识融合、加工,存储于Neo4j图数据库。在应用层分析中,筛选图数据库技术专利与应用专利节点,借助PKG子图划分和可视化,多维度归纳图数

数据库技术演化路径;运用 TextRank 算法提取领域关键词并扩展至 PKG 中,分析关键词年度变化趋势,同时

结合 PageRank 算法对应用领域节点影响力展开分析。

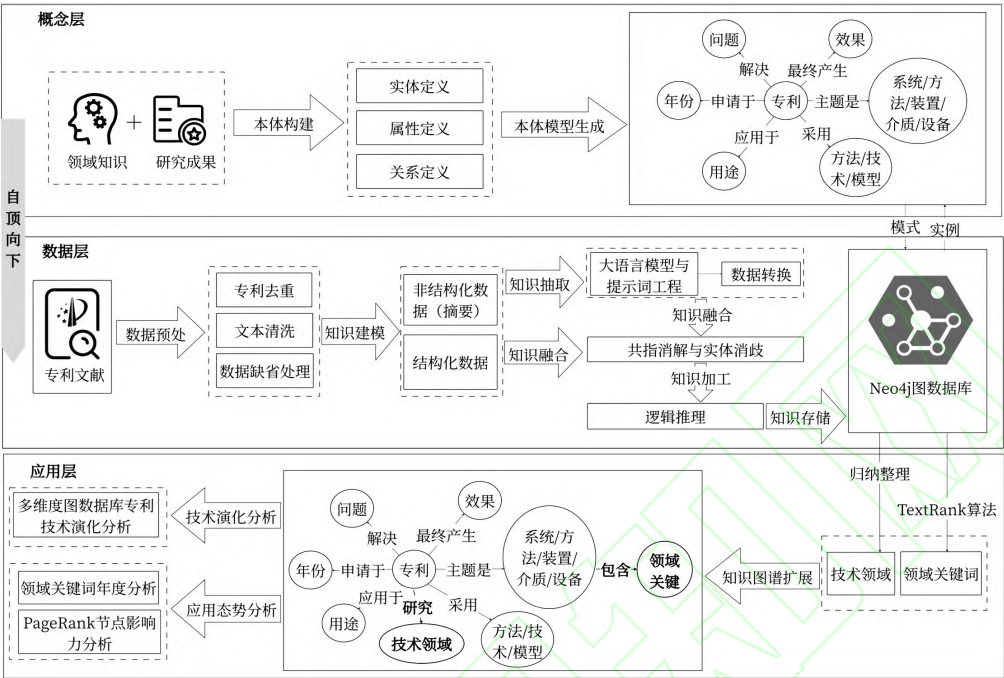


图 1 专利知识图谱构建及分析总体研究框架

3 图数据库专利知识图谱构建

3.1 模式层构建

在自顶向下的构建方法中,模式层的构建始于本体的定义与设计。已有研究从专利摘要中总结出系

统/方法/装置、应用领域、组件、方法、软件技术等实体类型^[7,32]。本文以此为依据,结合图数据库专利摘要的特点以及专利分析需求,采用手工构建本体的方法,归纳出 7 类实体和 6 类关系(见表 1 和表 2),这些元素普遍存在于摘要中,但并非所有摘要都包含。

表 1 图数据库领域 PKG 实体

实体名称	描述	实例
专利	专利本身,使用申请表表示	CN202310763174. 2
系统/方法/装置/介质/设备	专利的系统或方法或装置或介质或设备的名称	故障问题解决系统及方法
用途	专利整体的应用领域或用途	故障排查
方法/技术/模型	专利采用的方法或技术或模型	语义识别;实体抽取
问题	专利实际所解决的问题	故障排查准确性和效率低问题
效果	专利最终产生的效果	提高准确性和效率
申请年	专利申请的年份	2023

表 2 图数据库领域 PKG 实体关系

头实体	关系	尾实体	实例
专利	主题是	系统/方法/装置/介质/设备	CN202310763174. 2-主题是-故障问题解决系统及方法
专利	应用于	用途	CN202310763174. 2-应用于-故障排查
专利	采用	方法/技术/模型	CN202310763174. 2-采用-语义识别
专利	解决	问题	CN202310763174. 2-解决-故障排查准确性和效率低问题
专利	最终产生	效果	CN202310763174. 2-最终产生-降低沟通成本
专利	申请于	申请年	CN202310763174. 2-申请于-2023

3.2 数据层构建

3.2.1 数据获取与预处理

本文从 incoPat 专利数据库中检索“图数据库”领域的专利,检索式为:TIAB=" graph database" OR " graphDatabase" OR " graph databases" OR " graph-

Databases",检索日期为 2024 年 8 月 3 日,共获取 4448 条相关专利数据,选取中文专利名称、申请日和摘要等字段作为初始语料集。在数据预处理阶段,根据专利申请号、摘要和申请人信息进行去重,删除不相关或重复记录,使用同族专利的摘要替代少数缺失摘要的专

利,最终得到 3331 条专利数据。

3.2.2 知识抽取与数据转换

提示词工程指设计和优化输入提示词,以引导生成式 LLM 产生所需输出的过程,主要目的是通过精心设计的文本指令、问题、背景信息等,引导 LLM 产生准确、相关和有价值的回应^[33]。本文采用智谱 AI 的 GLM-4^[34] 系列 LLM,该系列模型支持长上下文、多模态、快速推理和多并发处理。研究利用 GLM-4-0520 模型和 Batch API,通过提示词工程对每个专利摘要要进行批量知识抽取,以获取特定类型的实体和关系。本文的提示词工程涵盖需求提示、示例提示、领域知识提示和输入输出提示,具体思路如图 2 所示。

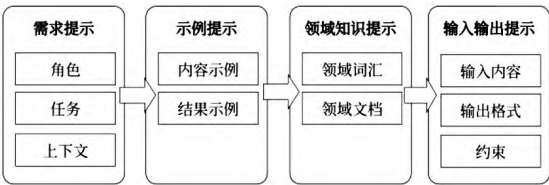


图 2 基于提示词工程的 LLM 抽取流程

需求提示包括以下几个方面:1)角色提示为 LLM

指定身份,以生成更符合情境和预期的输出^[35]。在本研究中,LLM 扮演“实体提取专家”,提升生成内容的专业性;2)任务提示描述任务指令,明确 LLM 的目标,是提示词工程的不可缺少的部分;3)上下文提示进一步解释提示语,帮助 LLM 更好地理解任务。示例提示指小样本提示^[36],使 LLM 通过少量实例进行小样本学习,达到与在大量数据集上训练的模型相当的高性能^[37]。示例的抽取结果以 JSON 格式输出,再转换为结构化数据并存储为 CSV 文件,用于知识融合、加工和存储。领域知识围绕“Graph Database”主题,整理中英文论文关键词,经翻译去重形成领域词汇,结合相关综述文献与行业报告构建本地知识库,依托智谱开放平台管理。通过检索知识库获取相关信息作为背景输入 LLM,辅助模型生成更准确、相关的答案。

输入输出提示包括“输入内容”即待抽取的专利摘要文本、输出格式的定义和约束条件设置。具体示例如表 3 所示,领域知识提示在执行抽取任务时通过接口调用本地数据库。

表 3 提示词示例

提示类别	提示示例
角色提示	你是一个实体提取专家
上下文提示	用户提供的文本是图数据库相关的专利摘要文本,‘系统/方法/装置/介质/设备’实体意味着专利的系统或方法或装置或介质或设备的名称,‘用途’实体表示专利整体是应用于哪个的领域或用途,‘方法/技术/模型’实体表示专利基于某些具体的方法或技术或模型,‘问题’实体表示专利实际解决了哪些问题,‘效果’实体表示专利最终产生的效果。这些实体及关系是专利摘要中普遍存在的,但非所有的,文本中没有提到的实体可以忽略
任务提示	你的任务是从专利摘要中提取实体构建专利知识图谱,三元组类型:专利-是-系统/方法/装置/介质/设备;专利-应用于-用途;专利-基于-方法/技术/模型;专利-解决-问题;专利-最终产生-效果
内容示例	‘本发明涉及一种基于语义识别的故障问题解决系统及方法,解决了故障排查过程费时费力,容易出错或遗漏信息导致故障解决工作的准确性和效率较低的问题,故障解决系统包括故障管理模块、数据映射模块和故障搜索引擎模块,解决故障的一般步骤为:构建动态更新的关系型故障数据库,生成故障信息表;对故障信息表进行实体信息抽取,建立实体关系;导入若干个实体关系,建立图形数据库;构建故障搜索引擎模块,进行故障匹配,获取故障问题解决方案。发明具有将故障问题形成闭环,降低故障排查过程中的沟通成本,故障处理高效快速、信息反馈及时、实用灵活便捷等优点’
结果示例	提取结果: ‘系统/方法/装置/介质/设备’:故障问题解决系统及方法’,‘用途’:[‘故障排查’],‘方法/技术/模型’:[‘语义识别’、‘实体抽取’],‘问题’:[‘故障排查准确性和效率低问题’],‘效果’:[‘提高准确性和效率’,‘降低沟通成本’]},
输入内容	文本内容:本公开提供一种智能问答处理方法、装置、设备及程序产品……(专利摘要内容,此处省略)
输出格式	提取结果只需返回以下格式的内容: ‘系统/方法/装置’:~,‘用途’:[],‘方法/技术/模型’:[],‘问题’:[],‘效果’:[]},
约束	不要复述问题,直接开始回答返回提取结果即可。文本中未提及实体直接设为空即可

3.2.3 知识融合与知识加工

知识融合指通过共指消解和实体消歧等方法,整合多源信息以构建统一的对象描述^[38]。本研究采用文本匹配、语义相似度分析和领域词汇参照的方法,对同类实体进行整理,例如将“自然语言处理”和“NLP 技术”统一为“自然语言处理技术”。知识加工是推理提取有用知识,支持后续应用。例如,针对专利的应用领域,抽取结果可能为“图数据库”或“图数据库查

询”。利用现有的实体、关系数据,可以推理“图数据库”的细粒度用途,促进对专利的深入分析。

3.2.4 知识存储

本研究通过 Neo4j-admin import 将数据批量导入 Neo4j 图数据库。经过“import. report”文件反馈数据集中的实体重复和缺失问题,快速检查数据集。最终得到 23265 个事实三元组,统计各类实体及关系的总数见表 4,PKG 的局部示意图如图 3 所示。

从“系统/方法/装置/介质/设备”视角(如表5):早期图数据库技术侧重在基础架构和查询优化,例如图数据库创建方法、子图分析和相似性查询,围绕基础构建和查询性能优化。2010–2015年间,专利重心转向扩展性和分布式处理,涵盖分布式图数据库创建、节

点部署、操作系统接口方法,以及索引管理和数据存储技术等。自2016年起,研究趋向更复杂的应用场景,涉及安全性、实时性及人工智能融合,如图数据库的加密方法、基于LLM的查询处理以及与机器学习技术结合。

表5 图数据库技术专利“系统/方法/装置/介质/设备”演化路径

2007–2009 年	2010–2015 年	2016–2024 年
一种使用子图和相似度查询分析、查询和挖掘图数据库的方法;创建图数据库的方法等	优化图数据库上的模式查询搜索的系统和方法;图数据库联机事务数据恢复的方法;对账处理的装置、程序和方法;生成和存储网络图数据库的技术;分布式索引及其并行生成方法等	全局列索引的方法,系统和计算机可读介质;用于处理图数据库的查询的系统;实体之间的关联性量化方法;员工不相容权限筛选方法、装置、计算机设备及存储介质;支持超大规模关系网络的图数据库构建方法;用于大规模图数据库系统的分布式存储和可视化的系统和方法;分布式原生图数据库中数据导出的方法和装置;图数据库与客户端之间的安全校验方法及系统等

从“方法/技术/模式”层面(如表6):早期专利聚焦基础图数据处理和表示,如闭包树和图元组编码。中期围绕模式查询优化、分布式技术和联机事务恢复,旨在提升性能、分布式处理能力及数据一致性,同时创新图数据索引与存储。后期,人工智能技术如机器学

习和自然语言处理等技术被引入,优化了图数据库查询、存储和交互等功能,在分布式环境中运用并发控制、设备集群、负载均衡等技术,通过路径遍历授权、加密信息嵌入和握手密钥校验等方法加强安全性和隐私保护。

表6 图数据库技术专利“方法/技术/模式”演化路径

2007–2009 年	2010–2015 年	2016–2024 年
闭包树;图形元组的编码与存储;显著性模型;随机漫步技术;节点间系数计算;查询语言和开放 API	图形查询分解;多查询执行引擎处理;随机漫步技术;有序键-值存储;子图查询;分布式信息处理技术;频繁子图挖掘;MapReduce 的并行处理;多线程执行;节点映射;最短路径算法;不完整的 Dijkstra 算法;深度优先搜索;加权路径计算等	图数据库模式验证;无锁哈希表的处理;边缘存储的访问方法;图嵌入向量的生产和查询;分片策略;云计算;半监督学习标签传播算法;机器学习模型分类器;卷积神经网络;压缩边缘存储;图数据格式化;元数据同步 Gremlin 语言;倒排索引;Brandes 算法;图划分;大语言模型;抗量子加密算法;版本控制;词向量模型;自注意力机制;数据快照;自动补全技术;估计成本模型;图嵌入距离预测;数据分区策略等

从“用途”、“问题”和“效果”分析(如表7):早期专利用于图数据库创建、查询和子图挖掘,解决数据存储结构优化和查询效率提升等基础问题,有效提高搜索效率,优化数据存储结构,奠定初级发展基础。中期探索分布式图数据库、事务处理和模式查询复杂的应用,解决异常操作导致的数据中断和不确定图数据子

图查询等问题,提升数据检索效率、安全性和可靠性,彰显技术稳定性与实用性。后期,相关专利用途扩展到多工作空间数据管理及图与关系数据库整合,大规模数据更新、UUID 冲突和查询效率低下等复杂问题,提升系统的综合性能,反映出图数据库技术在应对大数据挑战和多样化需求中的成熟。

表7 图数据库技术专利“用途”、“问题”、“效果”演化路径

实体维度	2007–2009 年	2010–2015 年	2016–2024 年
用途	图数据库查询;图数据库创建;图数据库子图挖掘;开放数据库系统;图数据库搜索查询	图数据库查询/检索/搜索查询;图数据库存储;图数据库划分;图数据库可视化;分布式图数据库;图数据库读取/数据访问;图数据库与关系数据库数据转换/整合;图数据库子图挖掘;关系图数据库系统等	图数据库查询/检索;图数据库存储;图数据库导入;图数据库更新;图数据库查询语言/语句;图数据库分析;图数据库与关系数据库数据转换/整合;耦合图数据库对管理;数据模型构建;图数据库测试;图数据库交互;图数据与其他数据库数据转化;RDF 图数据存储;分布式图数据库导入;关系图数据库看查询;图数据库安全查询;图数据库集群管理;图数据库权限管理等
问题	开销视图比较问题;主题图定义关系检索问题	区分不连接的子图关系问题;查询效率问题;异常操作事务导致的数据中断问题;生成关系图数据的方法问题;查询中属性的添加和检索问题;现有技术中协作计算机网络的不足;	图数据库节点数据容量瓶颈问题;超级节点和大图存储和处理难题;大规模数据集上数据查询速度慢的问题;大规模数据的定时更新问题;安全性差;分布式图数据可扩展性问题;待处理图数据的质量问题;图数据库存储分发的优化难度大;用户操作难度高;计算资源浪费问题;支持并发度低;大规模分布式图数据库集群效率较低的问题;图数据库计算过程中的内存占用问题等
效果	优化数据存储构;提高搜索效率;提高数据访问和更新效率等	提高查询和搜索性能;优化数据存储和管理;数据恢复和可用性;降低成本;提升数据访问性;数据安全和完整性;增强用户交互和灵活性等	图数据查询/检索;图数据提高查询和搜索性能;优化数据存储和管理;数据恢复和可用性;降低成本;提升数据访问性;数据安全和完整性;增强用户交互和灵活性等存储;图数据导入;图数据更新;图数据库查询语言/语句;图数据库分析;图数据库与关系数据库数据转换/整合;耦合图数据库对管理;数据模型构建;图数据库交互;图数据与其他数据库数据转化;RDF 图数据存储;分布式图数据库导入;关系图数据库查询;图数据库安全查询;图数据库集群管理;图数据库权限管理等

综上,图数据库技术从基础架构和查询优化起步,发展到分布式处理和扩展性增强,再到与人工智能融合及多样化应用。随着图数据库技术的成熟且应用范围扩大,其在未来数据处理和分析领域将发挥更为关键的作用。

4.3 图数据库应用态势分析

统计各年份“系统/方法/装置/介质/设备”实体包含各领域关键词的频率,构建 Cypher 查询语句: MATCH (p:专利)-[r:申请于]->(y:年份), (p)-[r2:主题是]->(s:系统/方法/装置/介质/设备), (s)-[:包含]->(k:领域关键词) RETURN y, year AS Year, k, name AS Field, COUNT(p) AS PatentCount。据查询结果绘制的权重前十的领域关键词年度热力图(图 7)显示,2010 年代中期起,图数据库应用研究呈增长趋势。综合 4.2 的分析可知,这得益于图数据库技术和性能不断成熟,如原生分布式图存储、图计算引擎的进步,及安全性和实时性提升;同时,数据复杂性增加推动图数据分析需求^[43],促进了图数据库应用发展。

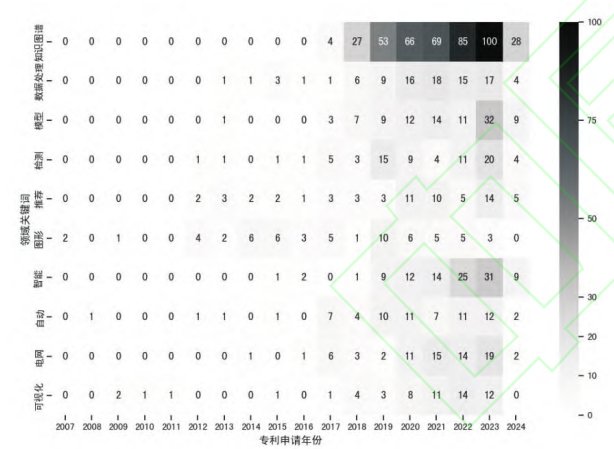


图 7 领域关键词年度热力图

其中权重前三的领域词汇为知识图谱、数据处理和模型。知识图谱应用包含构建、知识管理、搜索引擎优化、问答和推荐等,2017 年起,相关专利显著增长,成为图数据库应用热点。图数据库为知识图谱提供数据存储^[44],知识图谱则增强语义与推理能力,支持复杂知识表示^[45]。结合机器学习和自然语言处理,推动知识图谱自动化构建及在智能应用中的广泛应用,如 2018 年的软件缺陷知识图谱自动化构建方法和 2019 年的基于知识图谱的 Web 智能化测试方法。

“数据处理”相关专利出现于 2013 年,自 2018 年起进入快速增长阶段,涉及图数据的构建、存储、分析和可视化,如用图数据库优化电力网络建模以满足业务需求。“模型”相关专利通过构建、优化和应用模型解决实际问题,如构建风险传播模型评估信贷风险。2017 年起,“模型”专利缓慢增加,2023 年随着大模型

技术发展显著,如电力数据大模型构建、增强大语言模型在特定领域的问答能力,在这些应用中图数据库主要用于图数据存储、管理和分析。

PageRank 算法^[46]评估图谱中节点整体结构和中心性,节点得分反映其影响力和连接性^[47]。本文运行 PageRank 计算节点的影响力得分,表 8 列出得分前十节点,均为领域关键词节点。其中“知识图谱”得分最高,表示最具有中心性和影响力。“数据处理”在领域关键词中权重第二,但在此得分仅排第九,与其他领域关键词的连接性和交叉性较弱。“网络”和“知识”权重较低但其应用频率高且与其他关键词共现频繁因此得分较高,如图 8 为“网络”与其他领域关键词(白色节点)共现图。

表 8 PageRank 节点影响力得分

节点名称	得分
知识图谱	50.962
查询	9.790
模型	8.510
存储	8.415
智能	7.917
网络	7.589
知识	7.089
检测	6.770
数据处理	6.618
推荐	6.156

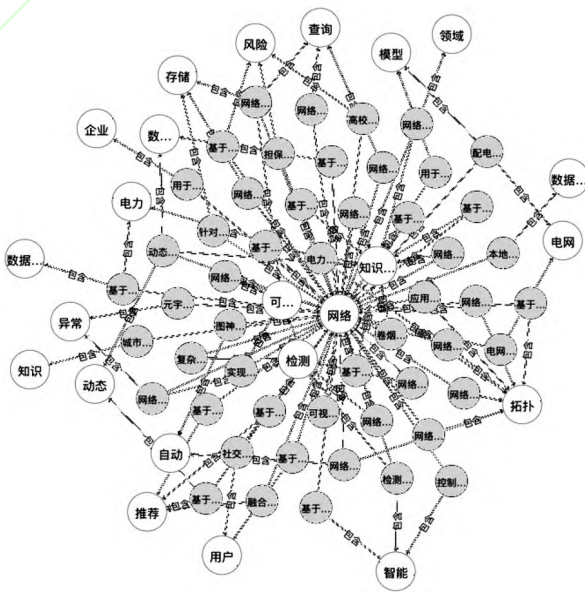


图 8 “网络”关键词与其他领域关键词共现图

5 结 语

本文以图数据库相关专利为研究对象,借助 LLM 对专利摘要进行知识抽取,构建图数据库领域 PKG。基于此,从方法/技术/模式、用途、问题、效果等维度剖析图数据库技术演化,并运用 TextRank 和 PageRank

算法挖掘应用领域关键词、分析节点影响力,实现应用态势分析。

研究表明,LLM 结合提示词工程,无需大量人工标注数据,就能高效完成知识抽取,解决了专利数据难以组织难题。PKG 可视化清晰展现图数据库技术发展脉络:从早期基础架构与查询优化,到中期分布式处理与扩展性提升,再到后期关注安全性、实时性及与人工智能深度融合,凸显其在存储、检索、稳定性及应对多样化需求方面的成熟与创新。TextRank 算法分析得到知识图谱、数据处理的模型等应用领域关键词。2017 年起,知识图谱相关专利显著增长,广泛用于构建、知识管理及智能应用;2018 年起,数据处理专利快速增长,聚焦图数据构建、存储与分析;2023 年,随着大模型技术发展,模型相关专利显著增加,并应用于解决实际问题。PageRank 算法中,“知识图谱”得分最高,中心性和影响力最强;“数据处理”连接性较弱;“网络”和“知识”等因高频共现得分较高。

本研究的局限性在于:仅运用 LLM 完成专利信息知识抽取,未将其融入 PKG 构建的知识融合加工、图谱扩展及专利分析环节。未来研究可利用 LLM 先验知识实现知识图谱补全,探索协同交互策略和优化模型微调更新机制,以增强模型对 PKG 结构与语义的理解及推理能力,助力专利挖掘、价值评估、布局规划等应用,推动专利领域智能化发展。

参考文献

- [1] 刘宇宁, 范冰冰. 图数据库发展综述[J]. 计算机系统应用, 2022, 31(8): 1-16.
- [2] Eldratt H A, Maatuk A M. Data migration from conventional databases into NoSQL: Methods and techniques [C]//2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA). IEEE, 2023: 370-375.
- [3] Das R, Soylyu M. A key review on graph data science: The power of graphs in scientific studies[J]. Chemometrics and Intelligent Laboratory Systems, 2023, 240: 104896.
- [4] 周贞云, 邱均平. 一门交叉学科的兴起:论图数据科学的构建[J]. 图书馆论坛, 2023, 43(4): 97-108.
- [5] 刘毅, 李岱素. 面向科技情报的专利分析研究综述及应用建议[J]. 科技管理研究, 2018, 38(22): 155-160.
- [6] 周贞云, 邱均平. 面向人工智能的我国知识图谱研究的分布特点与发展趋势[J]. 情报科学, 2022, 40(1): 184-192.
- [7] 张玲玲, 黄务兰. 基于 ChatGPT API 和提示词工程的专利知识图谱构建[J]. 情报杂志, 2025, 44(3): 180-187.
- [8] GUPTA A, RANGARAJAN K. Uncover this tech term: Transformers[J]. Korean Journal of Radiology, 2024, 25(1): 113.
- [9] 罗文, 王厚峰. 大语言模型评测综述[J]. 中文信息学报, 2024, 38(1): 1-23.
- [10] 梁佳, 张丽萍, 闫盛, 等. 基于大语言模型的命名实体识别研究进展[J]. 计算机科学与探索, 2024, 18(10): 2594-2615.
- [11] 于浏洋, 郭志刚, 陈刚, 等. 面向知识图谱构建的知识抽取技术综述[J]. 信息工程大学学报, 2020, 21(2): 227-235.
- [12] 丁建平, 李卫军, 刘雪洋, 等. 命名实体识别研究综述[J]. 计算机工程与科学, 2024, 46(7): 1296-1310.
- [13] 韦景竹, 祝培培, 张桐瑞. 服务科技创新策略的知识图谱构建[J]. 图书馆论坛, 2025, 45(4): 81-91.
- [14] 何喜军, 石安杰, 吴爽爽, 等. 基于知识图谱与强化学习的专利交易推荐研究[J]. 系统工程理论与实践, 2024, 44(10): 3330-3345.
- [15] 姜春涛. 自动标注中文专利的引文信息[J]. 现代图书情报技术, 2015, (10): 81-87.
- [16] 王密平, 王昊, 邓三鸿, 等. 基于 CRFs 的冶金领域中文专利术语抽取研究[J]. 现代图书情报技术, 2016, (6): 28-36.
- [17] 陈忆群, 周如旗, 朱蔚恒, 等. 挖掘专利知识实现关键词自动抽取[J]. 计算机研究与发展, 2016, 53(8): 1740-1752.
- [18] Wu H, Shen G Q, Lin X, et al. A transformer-based deep learning model for recognizing communication-oriented entities from patents of ICT in construction[J]. Automation in Construction, 2021, 125: 103608.
- [19] Huang Y, Feng X, Zhu M. Research on patent technology value evaluation method based on knowledge graph[C]//2022 5th International Conference on Data Science and Information Technology (DSIT). IEEE, 2022: 1-7.
- [20] Geng B. Open relation extraction in patent claims with a hybrid network[J]. Wireless Communications and Mobile Computing, 2021, 2021(1): 5547281.
- [21] 何玉, 张晓冬, 郑鑫. 基于 SpERT-Aggcn 模型的专利知识图谱构建研究[J]. 数据分析与知识发现, 2024, 8(1): 146-156.
- [22] 何喜军, 张佑, 孟雪, 等. 专利供需知识图谱半自动化构建及应用[J]. 情报杂志, 2023, 42(3): 139-150.
- [23] Gao Z, Feng A, Song X, et al. Target-dependent sentiment classification with BERT[J]. Ieee Access, 2019, 7: 154290-154299.
- [24] Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT[J]. Diagnostic and Interventional Imaging, 2023, 104(6): 269-274.
- [25] 李佳斌, 魏庭新, 曲维光, 等. 大语言模型下古诗笺注知识库的构建与应用[J]. 图书馆论坛, 2025, 45(3): 99-109.
- [26] 裴炳森, 李欣, 蒋章涛, 等. 基于大语言模型的公安专业小样本知识抽取方法研究[J]. 计算机科学与探索, 2024, 18(10): 2630-2642.
- [27] 赵建飞, 陈挺, 王小梅, 等. 基于大语言模型知识自蒸馏的无标注专利关键信息抽取[J]. 数据分析与知识发现, 2024, 8(Z1): 133-143.
- [28] 曾金晶, 何惠辉. 基于四螺旋理论的高校专利推荐知识图谱构建[J]. 情报探索, 2023, (10): 81-89.
- [29] Trappey A J C, Liang C P, Lin H J. Using machine learning language models to generate innovation knowledge graphs for pa-

- tent mining[J]. *Applied Sciences*, 2022, 12(19): 9818.
- [30] Lee J, Park S, Lee J. Exploring potential R&D collaboration partners using embedding of patent graph[J]. *Sustainability*, 2023, 15(20): 14724.
- [31] 王东清, 芦飞, 张炳会, 等. 大语言模型中提示词工程综述[J]. *计算机系统应用*, 2025, 34(1): 1-10.
- [32] 曹茹烨, 曹树金. 基于知识图谱的高价值专利技术创新演化研究[J]. *现代情报*, 2024, 44(6): 3-17.
- [33] Patil R, Heston T F, Bhuse V. Prompt engineering in healthcare[J]. *Electronics*, 2024, 13(15): 2961.
- [34] GLM T, Zeng A, Xu B, et al. Chatglm: A family of large language models from glm-130b to glm-4 all tools[J]. *arXiv preprint arXiv: 2406. 12793*, 2024.
- [35] 黄峻, 林飞, 杨静, 等. 生成式 AI 的大模型提示工程: 方法、现状与展望[J]. *智能科学与技术学报*, 2024, 6(2): 115-133.
- [36] 何丽, 曾骁勇, 刘杰, 等. 面向小样本命名实体识别的实体语义优先提示学习方法[J]. *计算机应用研究*, 2024, 41(12): 3622-3627.
- [37] Wang F L, Shi D, Aguilar J, et al. A few-shot learning method based on knowledge graph in large language models[J]. *International Journal of Data Science and Analytics*, 2024: 1-20.
- [38] 张晓华, 刘道伟, 李柏青, 等. 信息驱动的大电网运行态势知识图谱框架及构建模式研究[J]. *中国电机工程学报*, 2024, 44(11): 4167-4181.
- [39] 庞庆华, 董显蔚, 周斌, 等. 基于情感分析与 TextRank 的负面在线评论关键词抽取[J]. *情报科学*, 2022, 40(5): 111-117.
- [40] 张钰玲, 彭丽徽, 张艳丰, 等. 基于专利数据挖掘的我国智慧应急关联技术识别与发展趋势研究[J]. *情报科学*, 2023, 41(8): 139-146.
- [41] 李俊逸, 王卓, 马鹏玮. 图数据库技术发展趋势研究[J]. *信息通信技术与政策*, 2021, 47(5): 67-72.
- [42] Sagharichian M, Naderi H, Haghjoo M. Expregel: A new computational model for large - scale graph processing[J]. *Concurrency and Computation: Practice and Experience*, 2015, 27(17): 4954-4969.
- [43] 刘梦雅, 刘燕兵, 于静, 等. 图数据分析系统计算模型综述[J]. *计算机应用研究*, 2017, 34(11): 3204-3213.
- [44] 钱玲飞, 崔晓蕾. 基于数据增强的领域知识图谱构建方法研究[J]. *现代情报*, 2022, 42(3): 31-39.
- [45] 张云中, 祝蕊. 面向知识问答系统的图情学术领域知识图谱构建: 多源数据整合视角[J]. *情报科学*, 2021, 39(5): 115-123.
- [46] 宋敏, 张学人, 聂聪. 中美专利影响力研究: 基于 PageRank 算法[J]. *科学学研究*, 2024, 42(4): 721-732.
- [47] Park S, Lee W, Choe B, et al. A survey on personalized PageRank computation algorithms[J]. *IEEE Access*, 2019, 7: 163049-163062.