



开放科学
(资源服务)
标识码
(OSID)

金融领域标准文献知识图谱的构建与实现

赵伟 张览 望俊成

中国科学技术信息研究所 北京 100038

摘要: [目的/意义] 金融标准是实现金融交易信息系统互联互通、降低交易成本的基本条件,以系统的视角分析标准文献间的替代过程和相互关系等,将有助于推动标准情报的有效分析和金融领域标准化政策的提出。[方法/过程/结果] 本文采用基于RDM模型的知识图谱构建流程,构建了金融领域标准文献的知识图谱,实现并分析了金融领域标准—主题—关键词之间的相互关联关系,印证了RDM模型的可行性。通过上述技术和实践探索,以为金融领域标准化工作提供参考和借鉴。

关键词: 标准文献; 标准知识; 知识图谱; 金融领域

中图分类号: G350

The Construction and Realization of Knowledge Graph of Standard Literature in Financial Field

ZHAO Wei ZHANG Lan WANG Juncheng

Institute of Scientific and Technical Information of China, Beijing 100038, China

Abstract: [Purpose/Significance] Financial standards are the basic conditions for realizing the interconnection of financial transaction information systems and reducing transaction costs. Combating the replacement, iteration, cross reference and field formulation of standard documents from a global and systematic perspective will help support standard intelligence analysis and serve the formulation of standardized strategic planning in the financial field. [Methods/Processes/Results] This paper uses the knowledge graph construction process based on RDM model to build the knowledge graph of standard literature in the financial field, and realizes and analyzes the correlation between standards, topics and keywords in the financial field, which confirms the feasibility of RDM model. Through the above technical and practical exploration, it is expected to provide reference for the standardization work in the financial field.

Keywords: Standard literature; standard knowledge; knowledge graph; financial field

基金项目 中国科学技术信息研究所重点工作“金融大数据建设与知识服务(二期)——金融科技知识图谱构建”(ZD2020-03)。

作者简介 赵伟(1975-), 博士, 研究员, 研究方向为科技资源管理, E-mail: zhaowei@istic.ac.cn; 张览(1996-), 硕士研究生, 研究方向为产业竞争情报; 望俊成(1984-), 博士, 副研究员, 研究方向为科技政策与科技管理、文本数据可视化、大数据治理。

引用格式 赵伟, 张览, 望俊成. 金融领域标准文献知识图谱的构建与实现[J]. 情报工程, 2022, 8(6): 103-113.

引言

在大数据和数字经济迅猛发展的大背景下,知识图谱技术已成为科技文献资源服务的重要创新手段,面向标准领域的知识服务已经成为标准信息化发展的前沿热点问题^[1]。开展标准文献知识图谱构建和资源服务,有助于有序组织相关领域知识,揭示标准知识的整体关联性,让标准数据发挥更大效用^[2]。

然而,总体上国内外针对标准文献知识图谱的相关研究还十分有限。早期研究主要面向标准间的关系关联网络,如李景等^[3]采用文献计量学方法,基于相关行业的标准文献规模统计,对比分析国内外同行业标准文献数量的不同;李涛等^[4]通过建立标准文献网络,辨析标准文献间的引用关系和相似性,但该阶段的研究仍未达到标准内容知识元的粒度^[5-8]。而后又逐渐向标准知识关联领域开展探索^[9,10],如计雄飞^[11]基于对标准文献内容的分解,通过文本分类、特征提取、内容加工等建立了标准内容比对分析模型。随着语义网技术的推广,进一步面向语义网环境下标准知识关联的理论和方法开展积极探索^[1,12-17],如李景等^[18]通过获取和标注标准文本和网络信息,构建了实现语料编辑、语义标引、双语翻译功能的标准文献语料库;刘知民^[19]在开展标准的信息组织基础上,通过分词技术和信息挖掘技术形成了标准知识库。总体上,相关研究在对标准间关系的抽取深度和广度上仍有所欠缺,并将影响到知识图谱构建的质量和效用。赵伟等^[2]提出了用于描述标准知识的标准知识单元五元组,构建了面向标准文献的资源层—描述层—映射层的三层

知识图谱模型,提出了基于规则和LDA主题模型的标准文献知识图谱构建的集成式方法,为实践研究提供了有益的参考。

金融标准是实现金融交易信息系统互联互通、降低交易成本的基本条件,为金融相关部门实施有效管理提供了基础性支撑。金融标准通过文献载体将对企业、市场及产业发展产生重要影响,以系统视角分析标准文献间的替代过程、引用关系等,将推动标准情报的有效研究和标准化政策的形成与应用^[16]。因此,本文尝试构建金融领域标准文献的知识图谱,以期对金融领域标准化工作实践提供参考和借鉴。

1 金融领域数据的范围和特征

从分类角度,金融业包含了众多细分行业和领域,大类主要包括银行类、证券类、保险类等,细分领域可进一步分为债券、信托、货币、基金等资管计划、要素市场、征信贷款等。金融领域涉及到各行业的数据信息,具有数据量巨大、多维度和完备性等特征,这些信息又以文字、图形、表格等形式存储在大量文档中,格式非标准统一且呈碎片化存在,具有隐蔽性,因此采用集自然语言处理技术、处理非标准化数据和多层级关系挖掘技术,展现数据关联性于一体的知识图谱应用,成为了金融领域数据管理和利用的较好的支撑形式。

构建金融领域知识图谱,不仅将对多源异构数据进行知识化整合,从而为缺乏可解释性的机器学习算法提供有益借鉴,还可以通过建立金融行业海量多源异构数据的关联关系,挖掘数据的深层价值。构建金融领域知识图谱的

数据来源主要包括互联网上的舆情数据, 期货、股权、股票、上市公司等工商数据, 以及行研报告信息等。在金融知识图谱构建的已有应用研究中, 最为普遍的是用于信用评估、风控反欺诈、营销推荐和产业链分析等实践, 常见的实体包括人物、公司、地名, 常见的关系包括投资关系、竞争关系和股权关系等。但在实际相关行业环境里, 某种意义上几乎各行各业都与金融或多或少地相关, 如某国家标准(或战略文件)的发布, 可能影响银行信贷决策, 进而影响某地区企业群的融资规模, 最终影响相关上市公司的股票价格。我们希望通过构建知识图谱来支撑实现类似的关联分析。

2 标准文献知识图谱的构建模型

如前文所述, 赵伟等基于标准文献内容及

特征的分析, 分解了标准知识结构, 提出了标准知识单元五元组, 在此基础上设计了标准文献知识图谱 RDM 模型, 从资源层、描述层和映射层三个方面提出了标准知识图谱构建的理论模型。通过识别和挖掘标准之间的关联关系, 包括一般性题录关系、相互采用的知识关联关系和基于主题关联的多元关系, 充实和扩展标准文献知识关联的已有成果^[2]。标准文献中“范围”部分可视为论文文献中的“摘要”, 基于提取其中存在揭示主题内容的主题词信息, 挖掘标准实体和主题词之间的语义关系, 从而构建标准主题之间的关联性。

基于标准文献知识图谱 RDM 模型, 进一步建立了包括标准文献的知识抽取和知识链接两个步骤的集成方法^[1], 分别作用于不同的章节结构, 从而建立起标准文献知识图谱, 拓展实体关系类型。

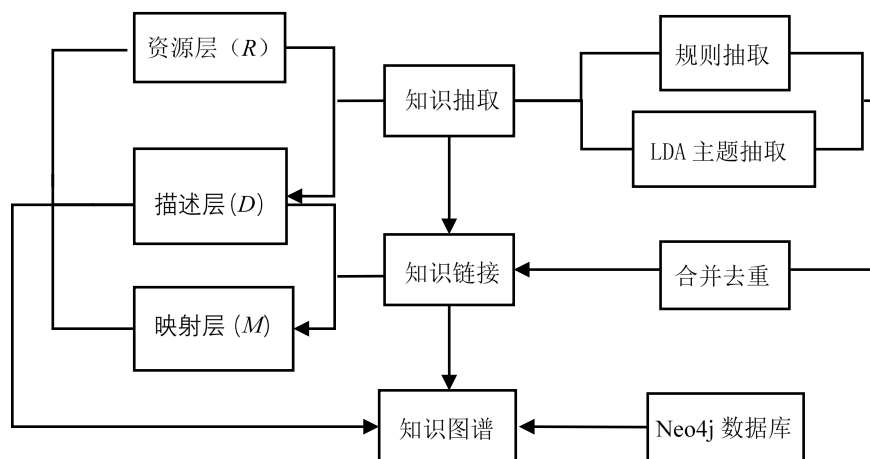


图 1 基于 RDM 模型的知识图谱构建流程

3 金融领域标准数据实体关系的抽取与主题词抽取

3.1 数据采集与预处理

(1) 实验环境

系统平台: Windows10 64 位;

处理器: Intel(R) Core(TM) i5-8250U, CPU
1.60GHz, 4 内核 8 逻辑处理器;

编程环境: Python3.8;

数据库: Neo4j-community-3.4.5。

(2) 数据获取与预处理

本文的研究数据采自中国标准服务网, 检索范围为现行标准, 不包括已废止的标准, 时间范围不限定。分别基于国际和中国标准分类号, 开展金融领域标准文献的遴选, 如表1所示, 从而获得73份国家标准和265份行业标准, 标准发布时间为1992—2020年。

表1 金融领域标准文献采集表

分类法	代码	名称
国际标准 分类号 (ICS)	03.060	金融、银行、货币体系、保险
	35.040	字符集和信息编码
	35.240.15	识别卡和有关装置
	35.240.40	信息技术在银行中的应用
	77.120.70	镭、钴及其合金
77.120.99	其他有色金属及其合金	
中国标准 分类号 (CCS)	A11	金融、保险
	A24	分类编码
	H15	贵金属机器合金分析方法
	H68	代表贵金属及其合金

采用自上而下的方法, 从文献文本中采集相关数据, 并分别利用不同的数据预处理技术对获取的实体数据进行加工。PDF文件是标准文献编辑生成后的主要形式, 本研究根据文献的段落格式并采用相关的信息抽取和逻辑结构转化技术, 将原始PDF文件转化为TXT格式或XML文档, 以便于更加清楚地显示标准文献的逻辑结构特点, 并开展人工核验和校对。

(3) 中文分词

在对标准文献范围文本内容进行中文分词的过程中, 需要先通过调用Python的jieba分词器, 来剔除标点符号以及非专业名词的外文词组; 其次将《金融行业主题词表》作为分词

组件的自定义领域词典, 经过不断迭代, 从而获得中文分词结果。

3.2 标准文献实体关系获取

基于针对标准文献前言、范围及规范性引用文件来开展知识抽取, 从而获得标准名称、发布年份、ICS、CCS、所属标准、代替标准、采用标准、采用方式、提出单位、归口单位、负责起草单位、参加起草单位、主要起草单位、起草单位、主要起草人、规范性引用文件和参考文献。进一步基于正则表达式规则抽取获得标准文献实体关系。

针对338篇标准文献进行显性知识抽取, 获得1534个实体节点和4716条关系, 如表2所示。

表2 显性知识抽取实体关系数量表

实体名称	实体数量 (个)	关系名称	关系数量 (条)
标准	338	—	—
所属标准	30	属于关系	110
采用标准	34	采用关系	34
代替标准	48	代替关系	48
规范性引用文件	625	引用关系	936
提出单位	31	提出关系	257
归口单位	8	归口关系	338
起草单位	420	起草单位 关系	2993
合计	1534	合计	4716

鉴于部分标准的原文较难获取, 面向互联网检索只能得到题录信息, 因此范围部分的文本数据量将少于检索标准的总量, 故本文将标准文献的标题合并到了范围数据中, 并基于

LDA 主题模型获取的主题抽取过程, 采用困惑度计算获得相对适合的主题数。设定的阈值将对困惑度的计算结果产生影响, 在阈值分别为 10、20、30 的条件下, 困惑度分别为 9、19、29, 进一步针对主题划分结果进行人工比对,

最终得到 19 类主题。经过统计合并关键词后共获得 136 个关键词, 从而形成 338 篇标准文献与主题类别的对应关系, 包括 289 条标准与主题的隐性关系和 460 条主题与关键词的隐性关系, 如表 3 所示。

表 3 金融领域标准文献主题抽取结果

类号	标准文献主题类别
1	银行; 金融机构; 业务; 规范; 注册; 接口; 产品; 客户; 市场; 经营; 系统; 描述; 内容; 基本; 方法; 数据交换; 结构; 元素; 证券期货; 设计; 证券公司; 统计; 共享; 信息交换; 银行业务
2	服务; 机构; 规范; 处理; 数据; 管理; 业务; 内容; 流程; 系统; 信息; 证券期货; 技术; 功能; 格式; 检测; 公司; 原则; 类型; 安全; 报送; 网络; 金融; 第三方; 基本
3	银行卡; 交易; 技术规范; 要素; 软件; 机构; 传输; 银行; 基本; 分类; 通用; 控制; 机制; 安全性; 信息; 密钥管理; 定义; 管理系统; 属性; 终端; 服务; 规范; 安全; 硬件; 支付
4	终端; IC卡; 应用; 受理; 规范; 金融; 借记; 贷记; 安全; 卡片; 发行; 设计; 开发; 管理; 集成电路; 系统; 制造; 单位; 集成; 维护; 研制; 部门; 功能; 支付; 硬件
5	管理; 技术; 密钥; 描述; 密钥管理; 环境; 保护; 信息; 数据; 密码; 交易; 安全; 生命周期; 原则; 加密; 银行业务; 数据元; 应用; 传输; 组织; 报文; 机制; 联机; 特定; 控制
6	分类; 部门; 活动; 统计; 报文; 编码; 个人; 代码; 统计数据; 规范; 范围; 特定; 传输; 公司; 指标; 银行; 方式; 应用; 数据交换; 方法; 客户; 硬件; 术语; 机构; 第三方
7	数据交换; 机构; 数据元; 协议; 交换; 系统; 格式; 结算; 报文; 市场; 内容; 交易; 接口; 业务; 数据; 文件; 规范; 共享; 定义; 信息交换; 联机; 数据字典; 证券期货; 应用; 传输
8	编码; 分类; 代码; 业务; 证券; 结构; 交易; 机构; 行业; 个人; 市场; 定义; 信息; 统计; 金融; 规则; 描述; 注册; 功能; 元素; 信息交换; 应用; 产品; 规范; 开发
9	支付; 安全; 产品; 接口; 机构; 技术规范; 个人; 金融; 保护; 系统; 技术; 评估; 平台; 结算; 应用; 内容; 传输; 移动支付; 市场; 证券期货; 定义; 功能; 清算; 检测机构; 注册
10	接口; 数据; 评估; 要素; 客户; 交换; 规则; 设计; 规范; 数据交换; 安全单元; 方法; 硬件; 证券公司; 范围; 借记; 移动; 编码; 单位; 金融业; 建设; 报文; 金融; 组织; 保险
11	金融; 应用; 设计; 接口; 支付; 文件; 规范; 移动支付; 系统; 维护; 平台; 管理; 开发; 集成; 研制; 发行; 内容; 制造; 受理; 结构; 要素; 应用规范; 产品; 单位; 行业
12	方法; 报文; 定义; 银行业务; 采用; 密码; 金融业; 传输; 保护; 编码; 加密; 密钥; 密钥管理; 金融服务; 规则; 描述; 产品; 环境; 原则; 范围; 数据字典; 元素; 方案; 金融机构; 个人
13	安全; 移动支付; 系统; 管理; 应用; 安全单元; 金融; 术语; 检测; 技术; 业务; 设备; 机构; 信息安全; 设计; 技术规范; 安全性; 定义; 规范; 指导; 评估; 金融服务; 基本; 行业; 开发
14	支付; 系统; 指标; 统计; 移动支付; 业务; 移动; 交易; 远程; 管理系统; 终端; 账户; 规则; 应用; 清算; 金融; 转接; 服务; 收单; 设计; 开发; 流程; 产品; 内容; 环境
15	数据; 受理; 信息系统; 格式; 业务; 内容; 传输; 安全; 类型; 收单; 属性; 行业; 密钥; 处理; 机构; 活动; 银行卡; 应用; 安全性; 证券; 元素; 单位; 框架; 账户; 保护
16	文档; 描述; 编制; 应用; 测试; 开发; 技术规范; 内容; 软件; 属性; 规范; 术语; 基础; 模型; 特定; 指南; 移动; 结构; 系统; 元素; 经营; 通用; 机构; 银行; 证券期货
17	风险; 保险; 指南; 评估; 证券公司; 控制; 基础; 方法; 内容; 制造; 经营; 规则; 服务; 框架; 单位; 金融业; 信息系统; 要素; 审计; 管理; 数据; 体系; 公司; 软件; 范围
18	SDMX; 统计数据; 元数据交换; 定义; 模型; 信息; 数据; 文档; 交换; 方案; 结构; 内容; 统计; 基础; 框架; 共享; 描述; 特定; 格式; 方法; 流程; 技术; 通用; 数据交换; 代码
19	信息系统; 证券期货; 机构; 审计; 保护; 指南; 行业; 管理; 运行; 基本; 经营; 服务; 规范; 信息安全; 建设; 维护; 公司; 测试; 评估; 市场; 证券公司; 指导; 结算; 金融业; 单位

4 金融领域标准文献知识图谱可视化的实现

本研究使用 Neo4j 创建知识图谱，首先将建立实体节点和实体之间的关系。采用 CQL “创建” 命令来创建无属性节点、创建属性节点、创建无属性节点之间的关系、创建属性节点之间的关系、创建一个或多个标签节点或关系。通过 CQL 语句，实现抽取出来的实体关系知识库与 Neo4j 图数据库之间的链接并推动知识图谱可视化。图 2 给出了标准文献知识图谱示意图，其中“金融业务”等红色节点代表初始标准文献，“中国人民

银行”等粉色节点代表标准提出单位，“全国金融标准化技术委员会”等黄色节点代表归口单位，“GB/T21078”等灰色节点代表初始标准所代替的标准文献，“银行业务证书管理”等绿色节点代表初始标准采用的标准文献。以实体节点 Standard 类型“金融服务生物特征识别第 1 部分：安全框架”为例，它与 JurisdictionOrg 类型“全国金融标准化技术委员会”之间的边关系为归口关系，与 AdoptStandard 类型“ISO 19092 1:2006《金融服务生物特征识别第 1 部分：安全框架》”节点之间形成了修改采用关系，展示了本文设计的核心实体与关系。

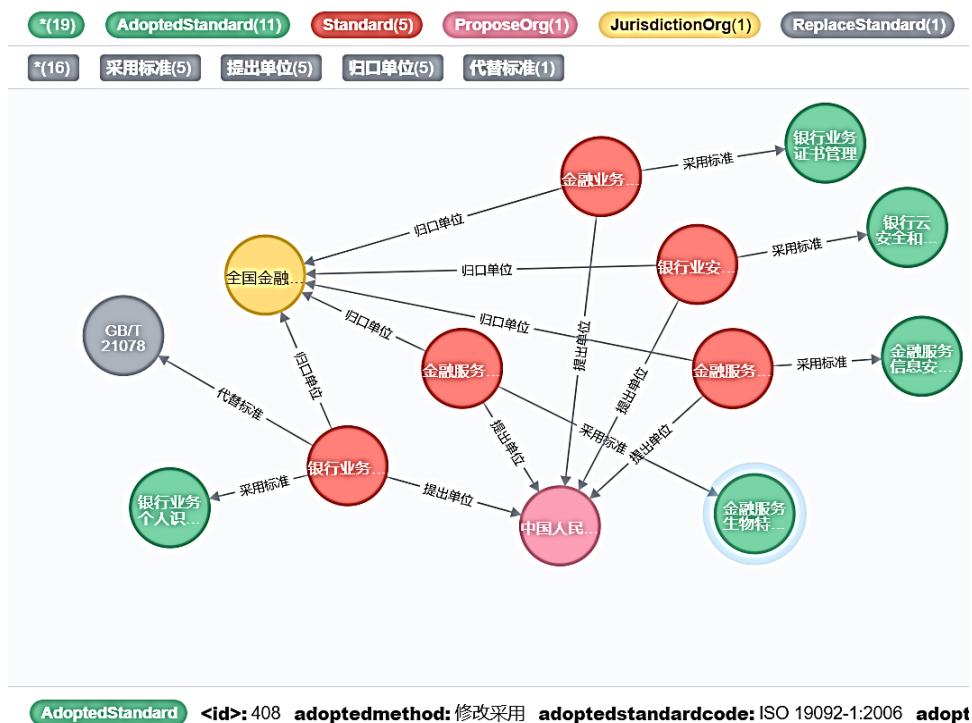


图 2 标准文献知识图谱示意图

(1) 标准文献中的归口单位关系。金融领域标准文献归口单位共 8 个，出现频次排在前 5 位的相关情况见表 4，图 3 显示了标准与归口单位之间的关系，其中绿色

节点代表标准文献，黄色节点代表归口单位，从实体节点的密集程度可以看出，最主要的归口单位是全国金融标准化技术委员会。

表 4 金融领域标准文献归口单位数量 Top5 排名

归口单位名称	出现频次
全国金融标准化技术委员会	292
全国金融标准化技术委员会保险分技术委员会	29
中国印钞造币总公司	4
中华人民共和国财政部	4
全国电子商务标准化技术委员会	3

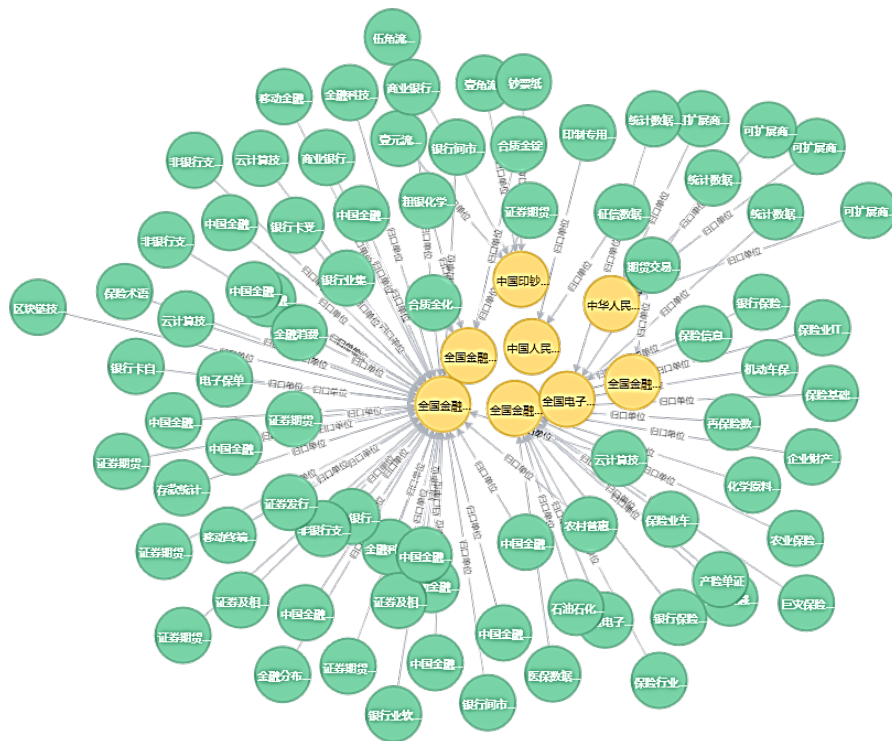


图 3 归口单位图谱片段

(2) 标准文献中的提出单位关系。基于本文数据范围，金融领域标准文献提出单位共 31 个，出现频次排在前 5 位的情况如表 5 所示。图 4 显示了标准与提出单位之间的关系，JR/T 0173—2020《银行业集中营运规范》、JR/T 0146.2—2016《证券期货业信息系统审计指南第 2 部分：期货交易所》、JR/T 0101—2013《银行业软件测试文档规范》等重要标准的提出单位均为全国金融标准化技术委员会。JR/T 0138—2016《银团贷款业务技术指南》、JR/T

0014—2005《银行信息化通用代码集》的提出单位均为国家开发银行。

表 5 金融领域标准文献提出单位数量 Top5 排名

提出单位名称	出现频次
中国人民银行	160
全国金融标准化技术委员会证券分技术委员会	31
全国金融标准化技术委员会	16
全国金融标准化技术委员会保险分技术委员会	6
中国证券监督管理委员会	5

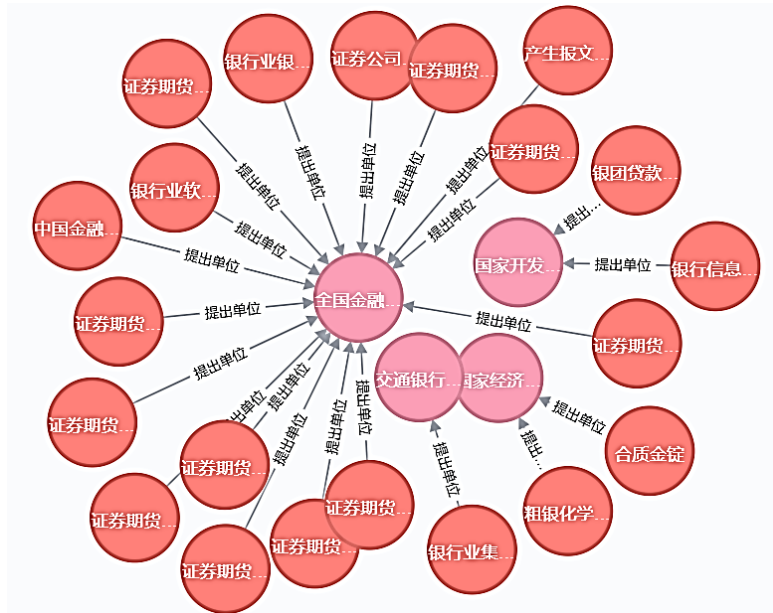


图4 提出单位图谱片段

(3) 标准文献中的采用关系。图5显示了标准与其他标准之间的采用联系，GB/T 27909.2—2011《银行业务密钥管理零售第2部分：对称密码及其密钥管理和生命周期》修改采用了ISO 11568.2:2005《银行业务密钥管理(零售)第2部分：对称密码系统及其密钥管理和生命周期》；GB/T 27911—2011《银行业安全和其他金融服务金融系统的安全框架》修改采用了ISO/TR 17944:2002《银行云安全和其他金融服务金融系统的安全框架》。

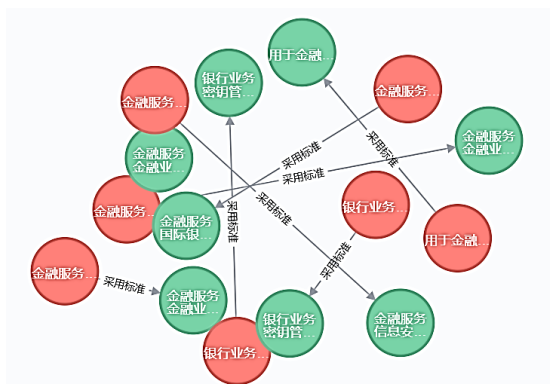


图5 采用标准图谱片段

(4) 标准文献中的替代关系。图6显示了标准与其他标准之间的替代关系。JR/T 0052—2020《银行卡卡片规范》替代了JR/T 0052—2009《银行卡卡片规范》，JR/T 0025.7—2018《中国金融集成电路IC卡规范第7部分借记/贷记应用安全规范》替代了JR/T 0025.8—2013《中国金融集成电路(IC)卡规范第8部分：与应用无关的非接触式规范》。

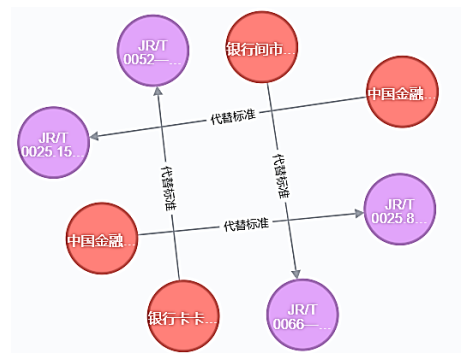


图6 代替标准图谱片段

(5) 标准文献中的起草关系。基于本文获取数据范围，金融领域标准文献起草单位的

共有 420 个，出现频次排在前 5 位的情况见表 6。图 7 以金融科技企业为例，显示了标准文献间通过起草单位构成的间接关系，其中红色节点代表标准文献，黄色节点代表起草单位。通过图谱可见，蚂蚁科技集团股份有限公司、京东数字科技控股股份有限公司、百度在线网络技术（北京）有限公司、深圳市腾讯计算机系统有限公司华为技术有限公司共同参与了 JR/T 0166—2020《云计算技术金融应用规范技术架构》、JR/T 0167—2020《云计算技术金融应用规范安全技术要求》、JR/T 0168—2020《云计

算技术金融应用规范容灾》、JR/T 0193—2020《区块链技术金融应用评估规则》等标准文献的起草。

表 6 金融领域标准文献起草单位数量 Top5 排名

起草单位名称	出现频次
中国金融电子化公司	145
中国工商银行	125
中国农业银行	109
中国建设银行	92
中国银行	89

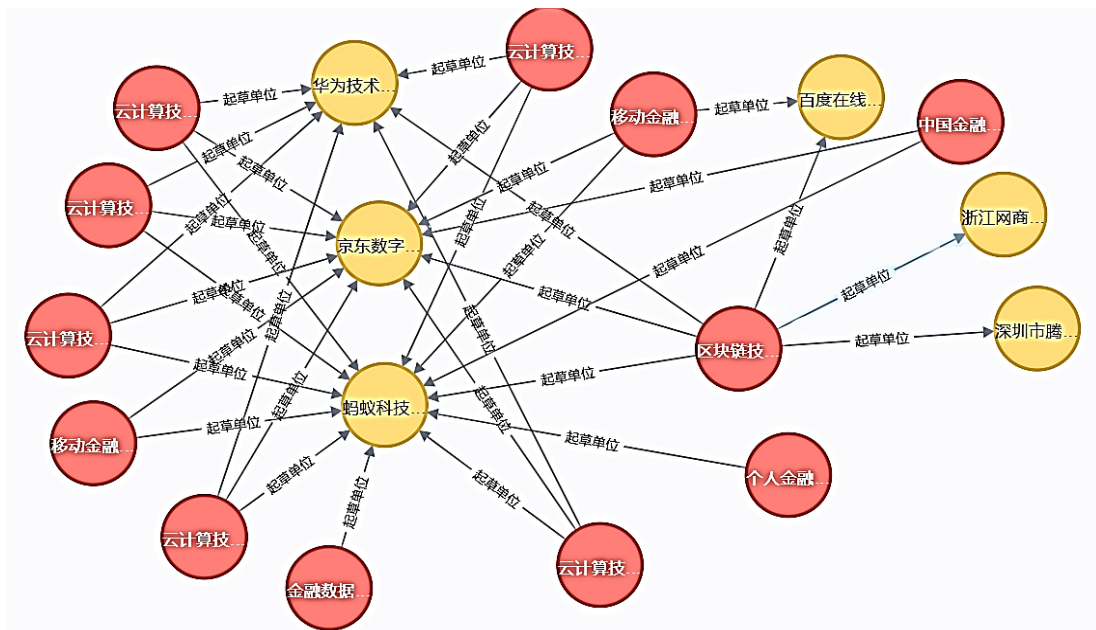


图 7 金融科技企业参与起草图谱片段

(6) 标准文献的主题关联关系。根据 LDA 主题模型抽取结果对我国现行金融领域标准文献进行划分，图 8 显示了不同标准文献间通过主题和关键词建立起来的链接关系，其中绿色节点代表标准文献，黄色节点代表关键词，蓝色节点代表主题，粉色节点代表起草单位。

在华为技术有限公司参与起草的标准文献中，有 6 篇标准文献涉及 4 类主题，分别是主题 1、主题 10、主题 12、主题 13。JR/T 0193—2020《区块链技术金融应用评估规则》与 JR/T 0168—2020《云计算技术金融应用规范容灾》同属于主题 10，该主题与 JR/T 0156—2017《移动终端

支付可信环境技术规范》所属的主题 13 通过关键词设计、规范、安全单元建立了主题关联关系。可见，基于 RDM 模型的主题关联图谱是可以

实现的，但图谱质量仍有待提升，有些关键词没有与之相连接的标准实体节点，这些都有待进一步完善。

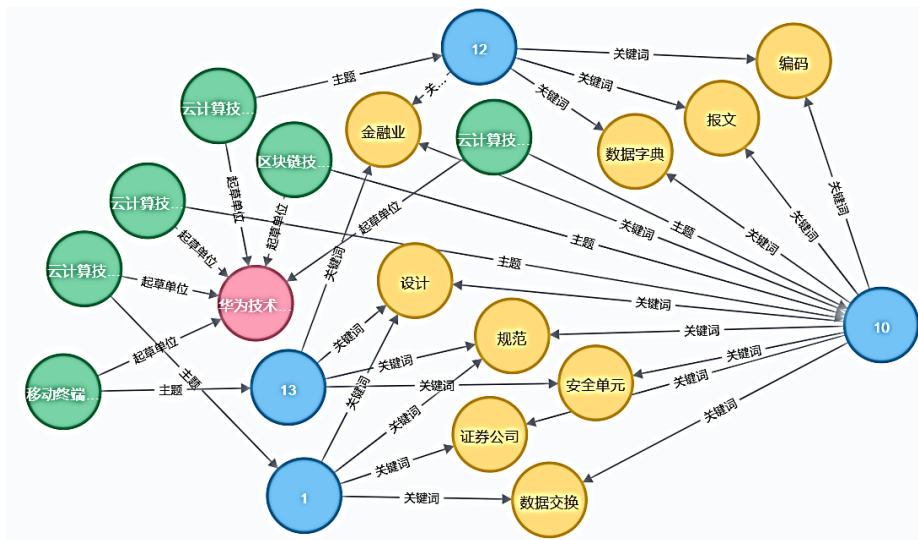


图 8 金融领域标准文献主题关联图谱 (片段)

7 结论

本文采用基于 RDM 模型的知识图谱构建流程，构建了金融领域标准文献的知识图谱。针对基于中国标准服务网获取到的金融行业标准文献，采取数据预处理、清洗加工、人工干预等手段，进一步开展显性知识和隐性知识的抽取，最后利用 Neo4j 数据库实现图谱的可视化并进行图谱解读。

图谱给出了标注归口管理、标准的主要提出单位等信息。除了基本信息查询，还通过 CQL 语句进行了深度查询，在揭示隐性关联的主题关联知识图谱中，实现了标准—主题—关键词之间的相互关联关系，印证了 RDM 模型的可行性。

但是，本文所构建的金融领域主题关联图

谱，但仍有一些无连接的关键词类型的节点，说明 LDA 主题模型的应用还有待进一步研究，标准文献的主题揭示效果有待进一步提升。

参考文献

- [1] 张览. 标准文献知识图谱的构建研究与实现 [D]. 北京: 中国科学技术信息研究所, 2020.
- [2] 赵伟, 张览, 望俊成. 标准文献知识图谱构建的模型设计与集成方法 [J]. 情报工程, 2021, 7(6):58-66.
- [3] 李景, 汪滨, 周洁, 等. 我国标准制定的领域动态趋势分析——基于国家标准馆 2006—2007 年度国内外标准文献新到馆藏 [J]. 图书情报工作, 2009, 53(1):56-60.
- [4] 李涛, 汪光阳. 标准相似性度量及其应用 [J]. 山西师范大学学报 (自然科学版), 2016, 30(4):29-33.
- [5] 刘华. 基于文献计量的国内外信息与文献国家标准对比研究 [J]. 图书情报工作, 2011, 55(12):51-55.
- [6] 王季云, 王宇. 基于社会网络分析法的标准网络测度指标研究——以五个主管部门国家标准网络比较为例 [J]. 中南财经政法大学学报, 2016(5): 21-

- 29.
- [7] 张君维, 马志远. 聚类方法在标准引用分析中的应用 [J]. 机械工业标准化与质量, 2016, 459(6): 40-42.
- [8] 张正敏, 郑东林, 孙冬梅. 基于引力模型的标准文献联系测度研究与应用 [J]. 标准科学, 2019(547):93-96.
- [9] 郭德华. 标准文献知识链接服务模式研究 [J]. 图书情报工作, 2011, 55(9):76-79.
- [10] 甘克勤, 丛超, 张宝林, 等. 基于划分的文本聚类算法在标准文献中的试验与对比研究 [J]. 标准科学, 2013(475):48-51.
- [11] 李涛, 汪光阳. 标准相似性度量及其应用 [J]. 山西师范大学学报(自然科学版), 2016, 30(4):29-33.
- [12] Noura M, Gyrard A, Heil S, et al. Automatic knowledge extraction to build semantic web of things applications[J]. IEEE Internet of Things Journal, 2019, 6(5):8447-8454.
- [13] 潘薇, 李景, 马志远. 语义网环境下的标准知识关联与服务探析 [J]. 标准科学, 2014(9):34-36+44.
- [14] 甘克勤, 马志远, 张明. 标准文献关联可视化研究与实践 [J]. 标准科学, 2015(1):34-38.
- [15] 梁薇, 马万钟, 王文君. 基于知识地图的标准知识可视化模型研究 [J]. 标准科学, 2015(7):62-64.
- [16] 李抵非, 田地, 胡雄伟. 基于深度学习的中文标准文献语言模型 [J]. 吉林大学学报, 2015, 45(2):596-599.
- [17] Kim S W, Gil J M. Research paper classification systems based on TF-IDF and LDA schemes[J]. Human-centric Computing and Information Sciences, 2019, 9(1):1-21.
- [18] 李景, 李国鹏, 汪滨. 标准文献语料库的构建与应用研究 [J]. 标准科学, 2013(5):6-11.
- [19] 胡扬, 闫宏飞, 陈翀. 面向金融知识图谱的实体和关系联合抽取算法 [J]. 重庆理工大学学报, 2020, 34(5):139-149.