

知识图谱在审计理论与实务教学中的应用研究

管 萍*, 王晓斌, 潘琥昊

上海工程技术大学管理学院, 上海

收稿日期: 2024年11月12日; 录用日期: 2024年12月11日; 发布日期: 2024年12月18日

摘 要

本文首先基于泛雅超星平台, 从知识图谱资源、知识关系、目标图谱以及问题图谱层面构建了审计理论通用知识图谱。其次, 基于Neo4j图数据库, 从审计数据获取、数据预处理、实体识别与关系处理、专家知识工程、自然语言理解、知识生成层面构建了审计实务领域知识图谱, 系统地阐释了审计课程的知识框架以及知识应用。

关键词

审计, 知识图谱, 领域知识图谱, 知识关系

Application Research of Knowledge Graph on Audit Theory and Practice Teaching

Ping Guan*, Xiaobin Wang, Huhao Pan

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Nov. 12th, 2024; accepted: Dec. 11th, 2024; published: Dec. 18th, 2024

Abstract

This article first constructs a general knowledge graph of audit theory based on the Chaoxing platform, from the perspectives of knowledge graph resources, knowledge relationships, target graphs, and problem graphs. Secondly, based on the Neo4j graph database, a knowledge graph in the field of audit practice was constructed from the aspects of audit data acquisition, data preprocessing, entity recognition and relationship processing, expert knowledge engineering, natural language understanding, and knowledge generation. The knowledge framework and knowledge application of audit courses were systematically explained.

*通讯作者。

Keywords

Audit, Knowledge Graph, Domain Knowledge Graph, Knowledge Relationship

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

知识图谱(Knowledge Graph)概念最早源自于语义网络,其基本要素是实体、关系、属性值,其中实体和属性值作为节点,属性和关系是边,节点之间通过边进行链接,形成网状的知识库,由此构成了知识图谱。学术界基于 CiteSpace 可视化软件进行知识图谱文献的研究已经非常成熟。除此之外,知识图谱在智能语义搜索、知识问答领域、数据分析决策等方面也在逐步应用[1]。在审计领域,由于审计知识涉及范围较广,审计知识之间的关系纷繁复杂,因此,将传统的审计方法与知识图谱相结合,开展多元异构数据深度挖掘和语义推理,提取三元组来构建审计知识图谱,方便审计信息使用者快速抓取信息,有助于审计决策的制定与实施[2]-[4]。本文基于理论教学和实务教学层面,首次尝试利用泛雅超星平台进行了审计理论教学层面知识图谱的构建,随后基于 Neo4j 图数据库研究了审计实务教学层面领域知识图谱的构建。

2. 审计理论知识图谱的构建

2.1. 建立知识图谱资源

将审计理论与实务课程的学习内容形成知识图谱,主要内容包括审计基本原理、审计测试流程、各类交易和账户余额的审计、对特殊事项的考虑、完成审计工作与出具审计报告、企业内部控制审计、质量管理、职业道德,涵盖了编制审计计划、收集审计证据,执行审计程序,形成审计报告的全流程。其中:基本理论内容有审计概述、审计计划、审计证据、审计抽样、信息技术对审计的影响、审计工作底稿;业务流程审计内容有风险评估和风险应对以及销售与收款循环、采购与付款循环、生产与存货循环以及货币资金循环。审计行业实践内容则包括审计报告、职业道德和独立性、内部控制等。

2.2. 构建知识关系

2.2.1. 父子关系

表示知识集合与知识点之间的关系,a是b的组成部分。例如审计职业道德作为知识集合,包括的知识点有诚信、客观公正、独立性、专业胜任能力和勤勉尽责、保密、良好的职业行为。

2.2.2. 前后置关系

表示两个知识点之间有先修后继关系,需要先学a后学b。例如审计目标知识点的学习在具体认定知识点之前。

2.2.3. 关联关系

表示两个知识点之间有相关性,即a和b有关联。例如内部控制测试和内部控制审计两个知识点具有较强的关联性。

基于泛雅超星平台的审计知识图谱资源和知识关系举例见表1。

Table 1. Audit theory knowledge graph
表 1. 审计理论知识图谱列表

知识图谱资源	主要知识关系
审计准则	父子关系：审计准则包含基本准则和具体准则 前后置关系：审计目标(前)和审计准则(后) 关联关系：审计准则与会计准则的关系
审计目标	父子关系：审计具体目标包含存在、完整性、权利和义务、准确性计价与分摊、分类与列报的余额认定；发生、完整性、准确性、截止、分类与列报的交易认定 前后置关系：审计总体目标(前)和审计具体目标(后) 关联关系：审计目标和审计准则的关系
审计风险	父子关系：审计风险包含重大错报风险，检查风险 前后置关系：重大错报风险(前)和检查风险(后) 关联关系：审计风险与重要性水平的关系
审计抽样	父子关系：审计抽样包含控制测试抽样和细节测试抽样 前后置关系：控制测试抽样(前)和细节测试抽样(后) 关联关系：抽样审计和全样本审计的关系
审计程序	父子关系：审计总体程序包含风险评估、控制测试和实质性程序。审计具体程序包含检查、观察、询问、函证、重新计算、重新执行、分析程序 前后置关系：重大错报风险评估(前)和重大错报风险应对的控制测试和实质性程序(后) 关联关系：审计程序与认定的关系
实质性测试	父子关系：实质性程序包含细节测试、实质性分析程序 前后置关系：控制测试(前)和实质性测试(后) 关联关系：控制测试和实质性测试的关系
审计报告	父子关系：审计报告意见包含无保留、保留、否定、无法发表审计意见 前后置关系：审计业务约定书(前)和审计报告(后) 关联关系：审计报告和审计工作底稿的关系。
内部控制审计	父子关系：财务报表审计包含内部控制测试；内部控制审计包含内部控制有效性的测试 前后置关系：控制测试(前)和内控审计(后) 关联关系：控制测试与内部控制审计的关系
职业道德	父子关系：审计职业道德道德包含诚信、客观公正、独立性、专业胜任能力和勤勉尽责、保密、良好的职业行为等 前后置关系：职业道德(前)和职业怀疑与判断(后) 关联关系：审计职业道德与会计职业道德的关系
智能审计	父子关系：大数据审计、人工智能审计、区块链审计等 前后置关系：传统审计(前)和智能审计(后) 关联关系：智能审计与智能财务的关系

2.3. 目标图谱

审计课程的目标主要包括掌握审计基本理论，熟悉审计业务流程，了解审计行业实践，培养学生具备职业道德、职业怀疑和职业判断的能力。将知识图谱任务点与课程目标建立关联。

2.4. 问题图谱

- 1) 疑难复杂问题：指向综合能力，要求学生发现审计知识之间的内在联系，重新排列组合概念及规

则, 进行选择、比较、价值判断。

2) 组合问题: 指向分析能力, 要求学生分析知识结构, 厘清审计概念之间关系, 最后得出结论。

3) 基本问题: 指向应用能力, 要求学生对基础知识记忆及理解, 并将审计知识应用于新的情境解决问题。

3. 审计实务领域知识图谱的构建

领域知识图谱由通用知识图谱演化而来, 知识获取、知识表示、知识管理、知识应用是领域知识图谱系统生命周期的四个重要环节。领域知识图谱构建的基本环节有数据获取、数据预处理、实体识别与关系处理、专家知识工程模块生成、自然语言理解、特征机器学习, 知识生成、图谱展示[5] [6]。

3.1. 数据获取

首先获取被审计单位会计记录以及内部控制手册、会议记录、分析报告、回函、盘点表等其他信息构建被审计单位审计证据数据库, 从财政部、注册会计师协会、内部审计师协会等官网上获取权威的审计相关法律法规和审计准则等构建审计制度库。

3.2. 数据预处理

数据预处理是进行数据清洗, 提高数据质量的关键步骤。会计记录等财务数据作为结构化的数据属性识别较为简单, 文本类半结构化数据和非结构化数据则需要经过段落识别、语句拆分、分词再进行属性识别, 形成审计术语文本语料库(类似审计词典)以及审计疑点或审计风险库。比如 Neo4j 图数据库中 CSV 就是使用文本文档形式存储的数据文件格式, 每一行代表一条数据, 每条记录包含数据实体、数据属性和数据关系。

3.3. 实体识别和关系处理

知识图谱的数据通常以三元组(S, P, O)的形式来表示, S 代表实体、P 代表关系、O 代表属性。财务报表审计语义网络中最主要的审计实体就是报表科目和报表项目, 如销售与收款循环业务活动中的营业收入项目和应收账款科目、采购与付款循环业务活动中的应付账款科目和一般费用项目、生产与存货循环业务活动中的存货项目以及货币资金循环业务活动中的银行存款和库存现金科目。内部控制审计的语义网络中最主要的审计实体则是内部控制。审计关系主要包括会计科目中上下级科目的父子关系、前后置关系, 业务数据与财务数据之间的相关关系, 企业内部数据和外部数据之间的相关关系, 通过实体识别和关系处理形成审计知识特征库。比如 Neo4j 图数据库中共设有两种节点类型, 分别是起始 Source 节点与目标 Target 节点, 节点与节点的关系则用 Relation 来表示。通过节点 - 关系 - 节点的方式, 进行领域知识图谱的构建。

3.4. 专家知识模块

专家知识系统一般来说是基于经验积累和案例分析基础上的。利用财务审计领域、计算机领域专家以及学术界、实务界专家的知识经验总结进一步拓展实体库, 并将审计可疑风险点与审计法律法规制度形成映射关系, 进一步来提高审计效率。

3.5. 自然语言理解与机器学习

自然语言理解模块是推荐、问答、搜索等系统的必备模块, 在文本信息处理系统中扮演着重要角色。从文本语料库中首先基于分词算法提取文本词汇, 筛选出特征项集合, 依据随机森林模型和决策树对词

频计算权重进行特征词提取，其次借助“词聚类组建”进行特征词聚类，基于词典、结构、语义信息并结合概念逻辑规则和相关库对特征词之间的相关关系进行计算。按照“聚类表示及组建”通过神经网络训练语言模型并采取分布式表示方法设置词向量，通过向量对比将距离较近的特征词予以聚类合并，文本空间视为规范化特征向量，特征词形成向量空间模型。通过自然语言理解与机器学习模块构建决策有用的审计术语文本语料库，提炼形成审计知识特征库。

3.6. 领域知识图谱生成与展示

经过数据获取、加工和处理形成审计领域概念框架，通过审计知识构建、知识关联建立、形成知识图谱。领域知识图谱区别于传统的知识搜索与推荐，知识的可视化强，决策相关性高，是未来大数据审计和智能审计发展的重要基础工具。基于 Neo4j 图数据库的审计领域知识图谱举例见表 2。

Table 2. Audit practice knowledge graph
表 2. 审计实务知识图谱列表

知识图谱环节	审计实务知识库
审计数据(获取)	被审计单位会计记录及其他信息库、会计审计法律法规制度库等
审计数据(处理)	审计术语文本语料库(或审计词典)、审计疑点(或审计风险)库等
审计实体	财务报表项目或科目等财务数据实体、内部控制等非财务数据实体等
审计关系	父子关系、前后置关系、相关关系等
审计规则(库)	会计审计类法律法规，审计准则、会计准则，专家经验、事务所经验等
审计问题(库)	收入舞弊、存货高估、商誉减值损失计提合理性等
审计案例(库)	近 5~10 年上市公司年报审计分析报告、会计造假案例等
审计知识特征(库)	报表审计、收支审计、经济责任审计、绩效审计、专项审计、内部控制审计的合法性、真实性和效益性决策指标等

4. 结语

大数据环境下，信息技术对于传统审计带来了各种变化及影响，知识图谱与审计的跨学科研究有助于审计信息化和智能化工作的研究推进。论文基于泛雅超星平台进行了审计理论知识图谱的构建，基于 NEO4j 进行了领域知识图谱的构建，在审计语义网络定义、审计数据管理、知识图谱生成方面做了有益的尝试，为知识图谱技术在审计学科中的应用研究提供了参考。

基金项目

2024 年校级教学建设项目课程知识图谱建设(k202403001《审计学》)。

参考文献

[1] 樊世昊. 基于知识图谱的审计方法研究[D]: [硕士学位论文]. 南京: 南京审计大学, 2018.

[2] 顾佳依, 管萍. 基于知识图谱的审计研究热点的文献计量分析[J]. 生产力研究, 2020(10): 157-160.

[3] 刘一鸣, 单鸿涛, 张高煜, 等. 审计知识图谱构建与推理在审计风险评估中的应用研究[J]. 计算机时代, 2023(3): 101-105.

[4] 王瑞萍, 刘峰, 杨媛琦, 等. 审计知识图谱的构建与研究——基于 Neo4j 的图谱技术[J]. 中国注册会计师, 2020(9): 109-113.

[5] 王倩玉. 领域知识图谱构建及其在财务审计的应用[D]: [硕士学位论文]. 南京: 南京审计大学, 2020.

[6] 徐玉芳. 高校财务审计领域知识图谱构建及应用[J]. 财会通讯, 2022(15): 118-122.