ICS 35.240.80
CCS C 07

T

团 体 标 准

T/CI 1125—2025

基于区块链技术的医疗健康知识图谱构建指南

Guide for the construction of medical and health knowledge graph based on blockchain technology

2025-08-01 发布　　2025-08-01 实施

中国国际科技促进会　发布
中国标准出版社　出版

目　　次

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。

本文件由中国科学院自动化研究所提出。

本文件由中国国际科技促进会归口。

本文件起草单位：中国科学院自动化研究所、东软集团股份有限公司、北京大学、中国医科大学附属第一医院、中山大学、广州大学。

本文件主要起草人：李明达、吴雅婧、王伟光、唐永强、毛文吉、吴中海、杨雪冰、顾正敏、黄方军、尹彦婷、沈晴霓、蒋彧琛、汤一凡、任至达、蔡巍、聂子皓。

引　言

本文件系统性地阐述了在区块链技术支撑下，构建医疗健康知识图谱的全周期方法论、关键阶段活动，以及相应的技术实现要求与最佳实践。其流程覆盖从严谨的数据源评估与可信接入、多元异构数据的规范化采集与深度预处理、基于先进技术的知识抽取与语义化、复杂的知识融合与冲突消解策略、面向应用优化的知识存储与多维表示，直至贯穿始终的知识验证、质量保障，以及基于区块链信任机制的图谱版本控制、发布与持续更新等核心环节。核心目标是在追求图谱知识准确性、完整性、一致性、时效性的同时，充分运用区块链技术对构建过程中的关键数字资产（如原始数据指纹、清洗规则、抽取模型、融合策略、审核记录、知识单元等）的权属、状态变迁和操作行为进行精确、不可篡改的记录与溯源，从而实现图谱构建过程的高度透明化、操作责任可追溯化，中间及最终成果的可信度增强。

本文件提出的构建方法视为与区块链平台要求紧密耦合、相互支撑的统一整体，共同构成基于区块链技术的医疗健康知识图谱构建的完整技术实施框架。必须强调，医学领域专业知识的深度融合及数据隐私保护与伦理合规的严格遵守，是贯穿整个构建过程的根本前提。

基于区块链技术的医疗健康知识图谱构建指南

1 范围

本文件提供了基于区块链技术的医疗健康知识图谱(以下简称“知识图谱”)构建的总则、数据采集与准备、数据清洗与预处理、知识抽取、知识融合、知识存储与表示、知识验证与更新的指导。

本文件适用于科研院所、医疗机构、第三方机构基于区块链技术对医疗健康知识图谱进行设计、开发等。其他基于区块链技术的知识图谱构建参照执行。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 5271.17 信息技术 词汇 第17部分:数据库

GB/T 15657 中医病证分类与代码

GB/T 15843.1 信息技术 安全技术 实体鉴别 第1部分:总则

GB/T 16751(所有部分) 中医临床诊疗术语

GB/T 17901.1 信息技术 安全技术 密钥管理 第1部分:框架

GB/T 17901.3 信息技术 安全技术 密钥管理 第3部分:采用非对称技术的机制

GB/T 20520 信息安全技术 公钥基础设施 时间戳规范

GB/T 22239 信息安全技术 网络安全等级保护基本要求

GB/T 25069 信息安全技术 术语

GB/T 30272 信息安全技术 公钥基础设施 标准符合性测评

GB/T 32905 信息安全技术 SM3密码杂凑算法

GB/T 32907 信息安全技术 SM4分组密码算法

GB/T 32915 信息安全技术 二元序列随机性检测方法

GB/T 32918(所有部分) 信息安全技术 SM2椭圆曲线公钥密码算法

GB/T 36344 信息技术 数据质量评价指标

GB/T 36626 信息安全技术 信息系统安全运维管理指南

GB/T 37092 信息安全技术 密码模块安全要求

GB/T 38625 信息安全技术 密码模块安全检测要求

GB/T 43572 区块链和分布式记账技术 术语

3 术语和定义

GB/T 5271.17、GB/T 25069、GB/T 43572界定的以及下列术语和定义适用于本文件。

3.1

医疗健康知识图谱 medical and health knowledge graph

以医疗实体为节点,以实体间语义关系为边,融合多源医疗数据构建的语义网络。

4 缩略语

下列缩略语适用于本文件。

ACID:原子性、一致性、隔离性、持久性(Atomicity,Consistency,Isolation,Durability)
API:应用程序编程接口(Application Programming Interface)
ATC:解剖学治疗学化学分类系统(Anatomical Therapeutic Chemical)
CPT:当前医疗程序术语(Current Procedural Terminology)
ChEBI:化学实体本体(Chemical Entities of Biological Interest)
DID:身份标识符(Decentralized Identifiers)
FDA:食品与药品管理局(Food and Drug Administration)
GDPR:通用数据保护条例(General Data Protection Regulation)
HGNC:人类基因命名委员会(HUGO Gene Nomenclature Committee)
HIPAA:健康保险流通与责任法案(Health Insurance Portability and Accountability Act)
ICD:国际疾病分类(International Classification of Diseases)
ICD-PCS:国际手术分类编码系统(International Classification of Diseases-Procedure Coding System)
ID:身份证标识号(Identity)
LOINC:逻辑观察标识符名称和代码(Logical Observation Identifiers Names and Codes)
MFA:多重要素验证(Multi Factor Authentication)
MeSH:医学主题词表(Medical Subject Headings)
NER:命名实体识别(Named Entity Recognition)
NLP:自然语言处理(Natural Language Processing)
OMIM:在线人类孟德尔遗传(Online Mendelian Inheritance in Man)
OWL:网络本体语言(Web Ontology Language)
PMID:PubMed唯一标识码(PubMed Identifier)
RDF:资源描述框架(Resource Description Framework)
SNOMED CT:医学系统命名法 临床术语(Systematized Nomenclature of Medicine—Clinical Terms)
UMLS:统一医学语言系统(Unified Medical Language System)
VC:可验证凭证(Verifiable Credentials)

5 总则

5.1 基础保障

5.1.1 制度建设

5.1.1.1 建立区块链医疗健康知识提供者的信息安全管理责任制,制定和公开管理规则和平台公约,落实真实身份信息认证制度,规避区块链信息安全风险。
5.1.1.2 定期开展合规性自查,留存审计报告上链。

5.1.2 技术保障

5.1.2.1 区块链框架符合GB/T 17901.1规定,运行环境符合GB/T 22239三级及以上的规定。
5.1.2.2 区块链采用密码算法,可选择的密码算法包括对称密码算法、非对称密码算法等,按GB/T

32905、GB/T 32907和GB/T 32918(所有部分)执行。技术钥管理包括但不限于对称密钥、非对称密钥、群密钥以及密钥派生等,符合GB/T 17901.3的规定。明确对称密钥更新频率、非对称密钥备份策略。

5.1.2.3 密码模块符合GB/T 37092二级及以上的规定,密码模块安全检测按GB/T 38625执行,公钥基础设施的标准符合性测评按GB/T 30272执行。

5.1.2.4 实体鉴别按GB/T 15843.1执行。

5.1.2.5 区块链中的可信时间源符合GB/T 20520的规定。

5.1.2.6 随机数生成和敏感安全参数生成按GB/T 37092执行,随机序列生成符合GB/T 32915的规定。

5.1.2.7 区块链运维管理中的身份认证与权限管理、密钥管理等工作的安全运维按GB/T 22239、GB/T 36626执行。

5.1.2.8 区块链功能性要求参见附录A。

5.2 知识图谱要求

5.2.1 知识图谱技术框架可参照GB/T 42131构建。

5.2.2 知识图谱的功能、性能、安全可参照T/CI 196执行。

6 数据采集与准备

6.1 数据源评估、许可、接入与注册

6.1.1 数据源评估

6.1.1.1 接入数据源之前进行评估,评估内容覆盖:

a) 数据质量:规范性、完整性、准确性、一致性、时效性、可访问性;
b) 数据内容:与知识图谱主题的相关性、可用字段、数据字典/元数据文档的可用性与质量;
c) 数据来源的合规性:数据持有权/使用权证明、获取数据的伦理审批文件、患者知情同意的范围与方式、是否符合相关数据保护法规;
d) 数据提供方的资质与信誉;
e) 数据敏感性分级:明确数据隐私保护等级;
f) 技术可行性:数据格式、接口类型与稳定性、更新频率、传输协议支持;
g) 成本与可持续性:获取成本、维护成本、长期合作可能性。

6.1.1.2 评估过程有文档记录,关键结论的摘要或哈希值可记录在链上作为决策依据。

6.1.2 数据共享协议与链上存证

6.1.2.1 与数据提供方签订的数据共享协议或数据使用许可。协议宜明确规定数据的使用目的、范围、期限、访问权限、保密义务、知识产权归属、数据销毁要求及违反约定的责任。

6.1.2.2 协议的关键条款摘要、全文哈希值、签署方数字签名、生效日期等信息在区块链上进行登记存证,形成可公开验证(对授权方)的法律约束基础。

6.1.2.3 涉及个人健康信息时,证据链能关联到有效的患者知情同意书或其哈希、授权凭证。

6.1.3 数据源链上注册

6.1.3.1 每个经过评估并获准接入的数据源,在区块链上完成一次注册操作。注册信息宜采用统一的数据模型,如预定义的智能合约结构或链下数据模式,包括但不限于:

a） 全局唯一的数据源标识符；
b） 来源机构的链上身份标识；
c） 数据类型与主题分类；
d） 数据格式描述；
e） 覆盖的时间范围与地理区域；
f） 6.1.1.1数据质量指标；
g） 更新频率与机制；
h） 数据负责人联系信息；
i） 指向链上存证的协议/许可记录的链接；
j） 技术接入点信息（如API端点描述，凭证不直接上链）；
k） 注册时间戳。

6.1.3.2 注册一经确认，即为该数据源在知识图谱生态中赋予了一个可信的数字身份和元数据档案。

6.1.4 接口配置与访问凭证管理

6.1.4.1 用于实际数据采集的API接口地址、认证凭证通过安全的带外渠道进行分发和存储，例如使用专门的密钥管理系统。

注：认证凭证如API Key、Secret、Token、证书等。

6.1.4.2 不宜将明文凭证硬编码在代码或配置文件中及上链。对特定系统或用户授予数据采集接口的访问权限时，授权记录的元数据可在链上进行登记，以便审计。授权记录的元数据包括被授权者身份、权限范围、有效期、授权时间。

6.2 数据采集、传输与完整性校验

6.2.1 数据采集

数据采集任务的执行遵守已签署的数据共享协议条款和所有适用的隐私保护法律法规。确保仅采集协议授权范围内且已获得必要同意（若涉及个人信息）的数据。采集脚本或程序有版本控制，其标识符可在执行日志中记录。

6.2.2 安全传输保障

数据从源系统传输到知识图谱构建的数据处理环境的过程中，全程使用强加密传输协议来保障数据的机密性和完整性。对于批量传输的大文件，可考虑使用文件级加密。

6.2.3 原始数据批次化与哈希锚定

6.2.3.1 采集的原始数据根据业务逻辑或处理效率需要，划分为有意义的批次或逻辑单元，如按天、按来源子系统、按数据类型等。

6.2.3.2 为批次原始数据或其压缩包、文件集合计算一个确定性的、抗碰撞的加密哈希值，将哈希值连同批次唯一标识符、关联的数据源链上ID、精确的采集时间戳、数据量统计、数据内容的简要描述等核心元数据，作为一个原子交易记录到区块链上。

6.2.4 任务日志上链

无论是定时调度还是手动触发，均为每次数据采集任务生成详细的执行日志，并将其关键信息记录上链。日志内容包括：

a） 任务唯一ID；
b） 计划执行时间与实际执行时间；
c） 触发者身份；
d） 目标数据源链上ID；
e） 尝试采集的数据范围描述，如时间段、特定表/主题；
f） 执行状态：成功/部分成功/失败；
g） 错误信息摘要（若失败）；
h） 成功采集到的数据批次的链上哈希引用列表；
i） 采集过程中使用的脚本/工具版本号。

6.2.5 数据质量扫描与报告存证

数据传输至处理环境后，宜立即进行一次数据质量扫描，检查数据的基本格式、完整性、一致性、分布特征等。扫描生成的质量报告摘要或其哈希值，可与对应的原始数据批次哈希关联并记录上链，并在链上进行存证，为后续的数据清洗提供输入。

7 数据清洗与预处理

7.1 处理操作

数据清洗与预处理操作包括但不限于：
a） 缺失值处理：基于医学知识、统计学方法或机器学习模型处理缺失数据，需明确记录所用策略及其依据；
b） 错误值识别与纠正：利用预定义的业务规则、医学常量范围、逻辑约束、校验码等发现并修正明显错误的数据；
c） 格式统一化：将日期、时间、度量衡单位、行政区划代码、证件号码等统一格式；
d） 医学术语映射：对文本中提及的疾病、症状、药品、检查、手术等术语，尝试映射到初步的内部或词典；
e） 数据类型转换：确保数据类型符合目标模式要求；
f） 冗余信息消除：识别并处理重复记录；
g） 异常值检测与处理：识别统计上的离群点，并根据业务进行修正、删除或标记；
h） 文本数据处理：对非结构化文本进行分句、分词、去除停用词、词干提取/词形还原、特殊符号处理等，为后续NLP任务做准备。

7.2 过程控制

7.2.1 工具控制

7.2.1.1 对用于执行清洗预处理任务的规则集、算法实现或可执行脚本进行版本控制管理。每个版本有清晰的文档说明其功能、参数、适用范围和预期效果。在应用到生产数据前，在测试数据集上进行充分的单元测试、集成测试和效果验证，确保其正确性和有效性。

7.2.1.2 经过验证的规则/脚本，或指向其在可信代码库中特定版本的唯一标识符、代码文件哈希、版本号在区块链上进行注册登记，创建不可篡改的规则/脚本库引用，后续执行清洗任务时在链上日志中明确引用所使用的规则/脚本的链上标识符及版本号。

7.2.2 任务日志上链

对每个或每批次数据执行清洗与预处理操作时，生成一条对应的原子性链上交易来记录该任务的详细信息，记录至少包括：

a） 唯一任务执行ID；
b） 执行主体身份（发起操作的用户或自动化服务的链上ID）；
c） 开始与结束时间戳；
d） 输入数据引用（指向链上记录的原始数据批次哈希）；
e） 输出数据集标识符及输出数据哈希；
f） 所使用的清洗规则/脚本的链上标识符及确切版本号；
g） 关键处理参数；
h） 执行环境信息摘要；
i） 执行结果状态（成功/失败/部分成功）及相关的量化指标，量化指标如处理记录数、改变字段数、删除记录数等。

7.2.3 过程记录

若在预处理阶段实施了数据脱敏、假名化或匿名化，则清晰记录所采用的具体技术方法、遵循的策略/规则集（其链上标识符）、操作范围及执行的操作。脱敏过程符合相关隐私法规要求且记录内容不泄露用于逆转脱敏的信息。

7.2.4 哈希存证

对经过清洗预处理后生成的每个新的数据集批次，计算其加密哈希值，并将该哈希值与对应的清洗任务链上日志记录进行强关联后一同记录上链。

7.2.5 数据质量评价与结果存证

清洗预处理步骤之后宜立即按GB/T 36344进行一次数据质量评价。使用预定义的质量维度和指标进行量化评估。生成的数据质量报告或其关键指标摘要、报告全文哈希与清洗后的数据批次哈希关联，并在链上进行存证。

8 知识抽取

8.1 抽取操作

知识抽取包括：

a） 命名实体识别：识别文本中提及的特定类别的实体，如疾病、症状、药物、检查、治疗、手术、解剖部位、基因、蛋白质等；
b） 关系抽取：识别并分类不同实体之间的语义关系，如“药物A【治疗】疾病B”“症状C【是】疾病D的表现”“检查E【用于诊断】疾病F”；
c） 事件抽取：识别文本中描述的特定事件及其参与者（实体）和属性，如识别一个“不良药物事件”，包括涉及的药物、发生的症状、患者信息、时间等要素；
d） 属性抽取：抽取实体的具体属性信息，如药物的剂量、用法、频次，疾病的分期、严重程度等；
e） 医学概念映射：将从文本中抽取的非标准、表述多样的实体映射到医学本体或术语集中的唯一

概念ID。

注：医学术语集如SNOMED CT、ICD、MeSH。

8.2 过程控制

8.2.1 工具控制

8.2.1.1 对用于执行知识抽取方法及所依赖的特征工程方法、词典资源、预训练模型、算法库、软件工具包进行版本控制。预训练模型需记录其训练相关信息，包括：

a) 所用训练数据集的描述，宜是数据集本身的哈希或链上引用；
b) 关键超参数配置；
c) 模型架构描述；
d) 在测试集上的性能评估指标。

8.2.1.2 模型、算法、规则库或指向其可信存储位置的唯一标识符、代码/模型文件哈希、确切版本号、以及上述相关的元数据和性能指标，宜在区块链上进行注册登记。

8.2.2 任务日志上链

对特定的预处理后数据批次执行知识抽取任务时，在区块链上记录详细的任务执行信息，至少包括：

a) 唯一任务执行ID；
b) 执行主体身份（用户或服务）；
c) 开始与结束时间戳；
d) 输入数据引用（指向链上记录的预处理后数据批次哈希）；
e) 所使用的知识抽取模型/算法/规则库的链上标识符及确切版本号；
f) 关键配置参数，如NER识别的实体类型列表、抽取的关系类型、概念链接的目标本体版本、置信度阈值设定等；
g) 执行结果状态（成功/失败）及摘要统计，如抽取的各类实体数量、关系三元组数量、事件记录数量；
h) 输出的知识单元集合的标识符或哈希引用。

8.2.3 哈希存证

抽取出的结构化知识进行批处理，对每批次抽取出的知识单元集合计算其整体内容的加密哈希值，并将该哈希值连同该批次的关键元数据一同记录在区块链上。

8.2.4 建立追溯链

从任何一个或一批被抽取出的知识单元，能准确地链接回生成的具体抽取任务执行记录，进而链接到该任务所使用的模型/规则版本，再链接到被处理的预处理后的数据批次，最终能追溯到最原始的数据采集批次及其数据源信息。

8.2.5 置信度与证据来源记录

8.2.5.1 为输出的每个知识单元提供一个置信度分数或置信度等级，并与知识单元本身或其批次哈希一同记录上链。

8.2.5.2 宜记录该知识单元在原始文本中的具体来源证据或至少是证据位置的指针/摘要。

注：来源证据如原文句子、段落、文档ID、文献PMID等。

9 知识融合

9.1 融合操作

融合操作包括：

a） 实体对齐/链接：识别并合并指向现实世界同一实体的不同表示，技术包括基于字符串相似度、属性相似度、网络结构相似度或预训练嵌入向量相似度的匹配算法等；

b） 关系融合/合并：处理关于同一对实体的相同或相似关系及处理相互矛盾的关系陈述；

c） 属性值融合：对于同一实体的同一属性，存在多个不同来源的值时，根据规则进行合并或选择；

d） 本体映射与对齐：当不同来源的知识基于不同的本体或模式时，建立映射关系；

e） 冲突检测与消解：主动识别知识库中存在的逻辑矛盾，并应用预定义策略或人工判断来解决。

注：优先级规则如权威指南＞临床路径＞文献证据。

9.2 过程控制

9.2.1 工具控制

对用于知识融合的具体策略、算法实现、规则集、本体映射文件、同义词词典、实体链接模型等进行版本控制。每个版本有明确的文档、测试和验证记录，并在区块链上进行注册登记，形成可信赖的“融合逻辑库”。

9.2.2 本体修改

9.2.2.1 对本体的任何修改，如新增类/属性/关系、修改约束、废弃元素，遵循变更管理流程，包括：

a） 变更提案；

b） 影响分析；

c） 专家评审；

d） 批准决策；

e） 实施部署。

9.2.2.2 每次本体的新版本及其变更日志进行版本标识，并将版本号、发布时间、变更摘要、本体文件哈希等信息记录在区块链上。

9.2.3 任务日志上链

执行每次或每阶段知识融合任务时，在区块链上记录详细的操作日志，包括：

a） 唯一任务执行ID；

b） 执行主体身份；

c） 开始与结束时间戳；

d） 输入知识单元集合的引用（指向链上记录的抽取知识批次哈希或之前的融合状态哈希）；

e） 输出的融合后知识状态的标识符或哈希引用；

f） 所采用的融合策略/规则/映射/本体版本的链上标识符；

g） 关键参数设置，如实体链接阈值、冲突解决策略选择；

h） 执行结果摘要，如处理的知识单元数量、新合并的实体数、解决的冲突数。

9.2.4 冲突及解决记录

融合过程中检测到的每个重要冲突及其最终的解决方式记录上链，包括：

a） 冲突标识符；
b） 冲突涉及的知识单元引用（指向链上记录）；
c） 冲突类型的描述，如事实矛盾、约束违反；
d） 采用的冲突消解规则/策略的链上标识符；
e） 人工裁决时，记录决策专家或委员会的链上身份标识、决策时间、决策结果及决策理由摘要或其哈希；
f） 最终被采纳或修改后的知识单元的引用。

9.2.5 哈希存证

在每次重要的融合操作完成后，对当前形成的知识图谱的一个稳定状态或其增量变化部分计算一个数据快照的加密哈希值，并将该哈希值与对应的融合任务链上日志记录进行强关联，并记录上链。

10 知识存储与表示

10.1 知识存储

10.1.1 选用能高效处理复杂图结构查询、支持大规模数据存储、并具备良好扩展性的链下图数据库系统。选择时考虑：

a） 图模型支持：使用RDF/OWL模型（适用于强语义表达和逻辑推理）或属性图模型（对节点和边的属性支持更灵活，更贴近某些应用场景），选择的数据库原生或高效支持所选模型；

注：常见的RDF存储有Apache Jena Fuseki、Virtuoso、GraphDB，常见的属性图数据库有Neo4j、JanusGraph、ArangoDB、NebulaGraph、TigerGraph等。

b） 查询语言：支持常见的图查询语言的程度和性能；
c） 性能与可扩展性：在预期的数据规模和查询负载下的读写性能、内存消耗、水平/垂直扩展能力；
d） 数据一致性与事务支持：对ACID事务或最终一致性的支持程度，是否满足应用需求；
e） 高可用与备份恢复：是否提供集群部署、自动故障转移及可靠的数据备份与恢复机制；
f） 安全性：是否提供细粒度的访问控制、加密存储、审计日志等安全特性；
g） 生态系统与社区支持：工具链、客户端库、社区活跃度、商业支持等；
h） 医疗合规性：存储系统及其部署环境满足HIPAA、GDPR、网络安全等级保护等对敏感数据存储的要求。

10.1.2 存储内容包括：

a） 知识图谱版本标识符与发布记录：全局唯一的版本号、发布时间戳、发布者身份、版本说明摘要或哈希；
b） 知识图谱快照完整性证明：对应链下存储的知识图谱数据在某个版本发布时的确定性哈希值；
c） 本体版本引用：该版本知识图谱所遵循的本体的链上标识符及版本号；
d） 构建过程溯源链锚点：指向构成此版本知识图谱的整个构建历史链条（从数据采集到最终融合验证）在区块链上记录的顶层索引或最终状态记录的指针/哈希；
e） 核心访问控制策略元数据：如定义访问权限的智能合约地址或策略版本哈希；
f） 关键统计与质量指标：如该版本包含的节点/边数量（按类型统计）、平均度数、关键质量评估指标摘要等；
g） 知识单元级溯源索引：对于需要极高可信度或细粒度追溯的场景，可考虑为每个或重要类型的知识单元在链上存储一个极简的索引记录，如其唯一ID、创建版本号、来源摘要哈希。

10.2 知识表示

10.2.1 明确并记录知识图谱内部采用的核心数据模型、选择的依据、模型选择的元数据信息，核心标识可在链上登记。

10.2.2 宜在知识图谱中使用国际或国内公认的医学本体和术语集来表示实体（节点）和关系（边）的类型以及属性值，包括但不限于：

a） 疾病与诊断：ICD、SNOMED CT、MeSH、罕见病本体（Orphanet）、疾病本体论（Disease Ontology）；
b） 药品与物质：RxNorm（药品标准化命名）、ATC、SNOMED CT、MeSH、ChEBI；
c） 症状与体征：SNOMED CT、MeSH、Human Phenotype Ontology（人类表型组数据库）；
d） 检查与检验：LOINC、SNOMED CT；
e） 手术与操作：CPT、ICD-PCS、SNOMED CT；
f） 解剖学：Foundational Model of Anatomy（解剖学基础模型）、SNOMED CT；
g） 基因与蛋白质：Gene Ontology（基因本体）、HGNC、UniProt（蛋白质数据库）；
h） 中医药：GB/T 15657、GB/T 16751（所有部分）；
i） 综合性本体/元知识库：UMLS。

10.3 过程控制

10.3.1 对选择和使用的标准/本体的名称、确切的版本号及在知识图谱中具体如何应用进行详细的文档化，其标识符和版本信息在区块链上进行注册登记。

10.3.2 遵循良好的知识图谱设计原则，包括：

a） 定义清晰的实体和关系类型层次结构；
b） 合理设计属性的粒度和数据类型；
c） 使用规范化的关系表达；
d） 考虑知识图谱的可扩展性和未来的查询需求。

10.3.3 建模决策过程和最终确定的本体/模式设计文档进行版本控制和归档。

11 知识验证与更新

11.1 知识验证与可信记录

11.1.1 知识质量与验证方法

定义清晰、可衡量的知识图谱质量，覆盖准确、完整性、一致性、时效性、可信度、相关性等维度，并制定包含多种互补方法的综合验证方法，方法宜文档化、版本化，且其标识符可在链上注册。验证方法包括：

a） 内部逻辑一致性校验：利用本体中定义的约束自动检测知识库中存在的逻辑矛盾，例如检查是否有记录显示某药物既能治疗又能引起同一种过敏反应；
b） 与外部金标准数据集的比对：将知识图谱中的部分知识与公认的、高质量的外部数据库或基准数据集进行比对，量化一致性；

注：基准数据集如FDA药品标签、OMIM遗传疾病库、ClinVar变异数据库。

c） 结构化的领域专家审阅：设计并实施由具备相关专业背景的医学专家参与的人工审阅流程，审阅基于明确的任务指引、统一的评价指标和易用的审阅工具，详细记录专家对知识单元的确认、质疑、修改建议或拒绝意见；
d） 基于文献证据的核查：对于从文献中抽取的知识，能追溯到原文出处，并由专家或自动化工具核

对其是否准确反映了原文信息及原文本身的证据等级；

e） 基于应用反馈的迭代验证：监控知识图谱在实际应用中的表现，收集用户反馈，特别是关于知识错误或不足的报告，将其作为持续验证和改进的重要输入。

11.1.2 任务日志上链

执行自动化验证程序时，执行记录宜上链，包括：

a） 唯一任务执行ID；

b） 执行时间；

c） 所用验证规则集/脚本的链上标识符；

d） 被验证的知识范围；

e） 验证结果摘要。

11.1.3 专家审阅过程与链上存证

11.1.3.1 专家审阅过程如下。

a） 专家身份认证与资质管理：参与审阅的专家关联其链上身份，专业资质和领域背景有记录，可使用VC。

b） 任务分配与追踪：记录审阅任务的分配。

c） 审阅操作的原子性记录上链：专家通过审阅工具提交的每一次审阅操作，都触发一次链上交易。对该交易进行记录，包括：

1） 审阅者链上ID，可假名化但能追溯；

2） 被审阅知识单元的链上标识符；

3） 审阅决策；

4） 给出的置信度评分（若适用）；

5） 审阅意见的文本摘要或其哈希，详细意见存链下；

6） 审阅时间戳。

11.1.3.2 为增强不可否认性，专家的数字签名与审阅记录绑定。

11.1.4 知识处置决策记录

基于自动验证发现的问题和根据专家审阅意见对知识单元进行的最终处置决策，连同决策依据、决策者身份、决策时间，都记录在区块链上，并与被处置的知识单元建立关联。

11.2 版本管理与可信发布机制

11.2.1 版本控制

11.2.1.1 建立清晰、规范的知识图谱版本控制机制，宜遵循语义化版本规范，即版本号格式为：

a） MAJOR：当做出不兼容的API更改或重大的本体/结构调整时递增；

b） MINOR：当以向后兼容的方式添加新功能、扩展知识覆盖范围或进行显著的内容更新时递增；

c） PATCH：当进行向后兼容的错误修复或小的知识修正时递增。

11.2.1.2 为每个版本定义明确的生命周期状态，如开发中、测试中、候选发布、已发布、已归档、已废弃。

11.2.2 版本发布链上登记

每个计划正式对外发布（供应用系统使用或公开）的知识图谱版本，在区块链上执行一次“发布登记”

交易。该交易是该版本知识图谱的“数字出生证明”,包含以下不可篡改的核心信息:

a) 全局唯一的、符合语义化规范的版本号;
b) 发布时间戳;
c) 发布操作的发起者/授权者链上身份;
d) 版本说明/变更日志的摘要,或指向详细文档的链接/哈希;
e) 对应的链下知识图谱数据快照的确定性哈希值;
f) 该版本所依赖的本体的链上标识符及版本号;
g) 指向构建此版本全过程链上记录链的最终锚点/顶层索引;
h) 适用时,关键质量指标快照,如节点数、边数、验证通过率等。

11.2.3 持续更新与迭代的可追溯性

对已发布知识图谱的任何后续更新,包括但不限于添加新知识、修正错误、适应新的医学进展等,都遵循第6章～第10章流程,形成新版本并在区块链上进行发布登记。区块链上的版本记录链清晰记录知识图谱随时间演化的路径。

11.2.4 历史版本查询与状态回溯

区块链上完整的版本发布记录和构建过程日志,支持授权用户查询任何一个历史发布版本的元数据、构建过程细节、当时的质量状况及对应的且通过哈希验证的链下数据快照。

11.2.5 知识弃用与撤回

发现已发布的知识存在严重错误或已过时,宜有机制将其标记为“弃用”或“撤回”,状态变更本身作为一种特殊的更新操作记录在区块链上,并关联到相应的知识单元或知识图谱版本。

附 录 A
（资料性）
区块链功能性要求

A.1 一般要求

A.1.1 区块链功能性包括身份管理与认证、访问控制、数据溯源与完整性保障、审计与监管支持。

A.1.2 身份管理与认证：建立全面、安全且适应医疗健康领域复杂参与方结构的身份管理和认证体系。该体系能唯一、可信地标识、验证并管理参与知识图谱构建、维护、治理和使用的所有实体身份。

A.1.3 访问控制：实现灵活、动态、策略驱动且能被区块链强制执行的访问控制模型，模型能基于已认证的身份及其关联的角色、属性，对知识图谱构建和应用全流程中的各类关键资源和操作实施精确到对象和动作级别的权限管理。

A.1.4 数据溯源与完整性保障：区块链的核心功能之一是作为不可篡改的分布式账本，为知识图谱构建过程中涉及的数据、元数据及操作活动提供可信的时间戳服务和持久化存证，构建端到端的溯源链条。

A.1.5 审计与监管支持：区块链平台设计为一个透明、高效且满足合规要求的审计底层，有力支撑内部质量控制、第三方审计以及来自卫生健康主管部门、数据保护机构等的监管活动。

A.2 身份管理与认证

A.2.1 身份类型与属性支持

A.2.1.1 能明确区分并管理多元化的身份类型，如医疗机构（数据提供方）、科研院所（算法提供方）、独立执业医师（知识审核专家）、患者组织（数据主体代表，若适用）、医药企业（应用方）、技术服务商（平台运维方）、监管机构（审计方）以及代表这些实体的个人用户或自动化服务/系统代理等。

A.2.1.2 支持为身份附加可验证的属性，如机构资质证书哈希、医师执业资格认证信息、专家领域标签，可考虑采用W3C DID及其关联的VC模型增强身份的可信度和互操作性。

A.2.2 标识符规范与生命周期管理

为每个实体分配一个在区块链网络内全局唯一、防篡改且生命周期内稳定的标识符，明确标识符的生成、注册、解析、更新、恢复及在实体退出或失信时的吊销/冻结机制。安全记录身份生命周期管理过程中的关键操作，如注册审批、信息变更、状态转换，操作日志具备不可篡改性，关键变更事件宜上链存证。

A.2.3 强认证机制与集成

A.2.3.1 若采用符合公私钥密码体系（如国密SM2、ECDSA）进行数字签名验证，提供基于强密码学的认证方法。

A.2.3.2 宜采用MFA策略以提升安全性。

A.2.3.3 具备与现有可信身份认证基础设施（如国家/行业PKI体系、OAuth/OpenID Connect服务、机构内部身份管理系统）安全集成的能力，实现身份的联邦认证或映射。

A.2.4 角色与权限关联

身份管理系统与访问控制模块紧密集成，支持将具体的角色或权限精细化地绑定到已认证的身份上。

A.3 访问控制

A.3.1 访问控制模型

以基于角色的访问控制为基础，同时结合基于属性的访问控制以应对医疗场景下复杂的权限决策逻辑。结合基于属性的访问控制策略可利用诸如用户属性（如角色、部门、资质认证）、资源属性（如数据敏感等级、知识领域、来源机构）、环境属性（如访问时间、地理位置、IP地址）以及操作类型等多种因素进行动态授权判断。

A.3.2 策略定义、存储与管理

提供清晰的方式定义访问控制策略，如使用策略语言或通过智能合约编码。策略本身进行版本控制，记录创建、修改、审批、激活、停用等管理操作，策略的关键版本或哈希值宜上链存储或锚定，确保策略本身的完整性和可追溯性。

A.3.3 强制执行与审计

A.3.3.1 访问控制决策逻辑嵌入到区块链的执行层（如通过智能合约的修饰符或内置访问控制逻辑）或通过可信的链下强制执行点（如API网关结合链上策略验证）来保障。

A.3.3.2 可靠记录所有对受控资源或操作的访问尝试，无论成功与否，记录的关键信息包括请求者身份、目标资源标识、请求的操作类型、时间戳、访问结果、执行节点信息等。授权成功的操作记录上链存储，失败访问尝试的日志记录宜安全存储并可供审计。

A.3.4 权限查询与审计接口

提供接口，允许授权管理员查询任何特定身份当前的有效权限集及查询权限分配、变更的历史记录。审计人员能便捷地查询和分析访问控制日志。

A.4 数据溯源与完整性保障

A.4.1 元数据上链策略

A.4.1.1 明确定义上链的关键元数据，以确保溯源的有效性。除基础的来源、时间戳、哈希值外，考虑包括数据所有权/管理权信息、数据使用许可/同意状态的证明（或其引用）、详细的处理步骤描述（如清洗规则ID、转换函数签名）、模型训练相关信息（数据集哈希、超参数摘要）、知识抽取置信度分数、人工审核意见的数字签名、知识图谱本体版本链接、相关联的业务交易ID等。

A.4.1.2 元数据上链的粒度根据业务需求确定，能保证能追溯到数据处理批次和关键操作步骤，理想情况下能追溯到更细粒度的单元（如单个知识三元组的来源摘要）。

A.4.2 操作日志结构化与关联

A.4.2.1 以结构化的方式将记录预定义的关键构建操作（数据引入、预处理、特征工程、模型训练/调用、知识抽取、融合、验证、发布等）的执行记录上链。

A.4.2.2 日志条目包括明确的事件类型、执行主体身份、时间戳、对输入（如数据批次哈希、模型版本ID）和输出（如结果哈希、新知识ID范围）的引用、使用的具体参数或配置信息、以及执行状态。链上日志记录通过内在逻辑（如交易输入/输出关系、显式引用）或链下索引辅助，形成一个可导航的、有向无环图

式的溯源路径，允许从最终的知识图谱实体反向追溯其完整的“制造”历史。

A.4.3 不可篡改性、不可否认性与时序性

利用区块链的哈希链结构、共识机制保证上链记录不被篡改；结合数字签名技术确保操作记录由声称的身份发起且不可否认；区块链固有的区块顺序和时间戳机制为所有记录提供了严格的时序证明。

A.4.4 可验证性机制与工具

提供明确的机制和配套工具/API，允许授权方（如数据所有者、审核员、监管者）便捷地查询链上溯源信息，并能执行链上链下一致性校验，如给定一个链下数据文件或知识图谱快照，工具能计算其哈希值并与链上记录的对应哈希进行比对，以验证其在特定时间点的完整性和未被篡改。

A.5 审计与监管支持

A.5.1 线索审计

完整、准确、及时地在链上账本中记录所有对系统安全、数据完整性、业务流程正确性有影响的关键事件（如身份管理操作、权限策略变更、数据上传/下载/删除元数据记录、模型部署/更新、关键算法参数修改、审核决策提交、知识图谱版本发布/回滚、敏感数据访问尝试等）都按照预定义的、结构化的格式（如包括事件ID、时间戳、用户ID、活动描述、源IP、结果状态等字段）。审计日志的格式宜参考相关安全标准或行业实践。

A.5.2 查询与分析能力

A.5.2.1 提供API和/或可视化工具，允许授权审计人员能够根据时间范围、事件类型、操作者身份、关联对象标识符、地理位置等多个维度进行灵活、高效的审计日志检索、过滤、聚合和关联分析。

A.5.2.2 支持生成符合审计要求的报告。

A.5.2.3 考虑提供订阅机制，允许审计系统实时接收特定类型的链上事件通知。

A.5.3 监管节点与接口支持

A.5.3.1 具备支持监管机构作为特殊类型的节点（如观察员节点、审计节点）安全接入网络的能力。此类节点拥有对特定范围或全部交易数据的只读访问权限，以便进行实时或准实时的合规监控、数据抽查和风险评估。

A.5.3.2 提供符合监管要求的数据接口，如数据报送接口。

A.5.4 审计日志生命周期管理

制定审计日志的存储策略，包括链上存储期限（考虑到区块链存储成本和无限增长问题，可能需要策略性地将极旧的、非核心的日志摘要化或归档至链下，但保留链上索引和完整性证明）及链下归档日志的安全存储、访问控制和销毁策略，确保满足法规对审计记录保存年限的要求。

参 考 文 献

[1] GB/T 42131 人工智能 知识图谱技术框架

[2] T/CI 196 医疗知识图谱构建技术要求

[3] 中医药区块链知识图谱白皮书(2023版)

[4] 中国信息通信研究院,浙江大学医学院附属邵逸夫医院. 卫生健康领域区块链应用白皮书[R]. 杭州:中国卫生健康科技创新与学科建设大会,2024.

中国国际科技促进会
团 体 标 准
基于区块链技术的医疗健康知识图谱构建指南
T/CI 1125—2025
*
中 国 标 准 出 版 社 出 版 发 行
北京市朝阳区和平里西街甲2号(100029)
网址 www.spc.net.cn
总编室:(010)68533533 发行中心:(010)51780238
读者服务部:(010)68523946
中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销
*
开本 880×1230 1/16 印张 1.5 字数 15 千字
2025年11月第1版 2025年11月第1次印刷
*
书号:155066·5-18258 定价 49.00 元

T/CI 1125-2025