|  |  |
| --- | --- |
| ICS  | 35.240.40 |
| CCS  |

|  |
| --- |
| D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png CCUA |

L67 |

中国计算机用户协会团体标准

T/CCUA LX003—2020

不良资产管理领域知识图谱 技术要求

Knowledge graph for asset management——Technical requirement

（本草案完成时间：2021年12月29日）

XXXX - XX - XX发布

XXXX - XX - XX实施

中国计算机用户协会  发布

目次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 2

5 概况 3

5.1 领域相关特点 3

5.2 领域相关限制 3

6 知识图谱技术架构 3

7 知识图谱构建和应用 4

7.1 源数据 4

7.2 数据处理 4

7.3 知识抽取 4

7.4 知识表示 4

7.5 知识融合 5

7.6 知识存储 5

7.7 图谱中心 5

7.8 知识推理 6

7.9 知识图谱应用 6

8 知识图谱系统运维 6

9 安全 7

参考文献 8

1. 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。

请注意本文件的相关内容不涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国计算机用户协会云应用分会和华融融通（北京）科技有限公司共同提出。

本文件由中国计算机用户协会归口。

本文件起草单位：中国计算机用户协会云应用分会、中国华融资产管理股份有限公司，北京邮电大学，北京航空航天大学，中国信达资产管理股份有限公司，华融融通（北京）科技有限公司，华鸿汇德（北京）信息技术有限公司。

本文件主要起草人：唐常芳、傅湘玲、王宝会、彭雷、袁佳宁、张昕夏、黄笑童、赵蒙、郑艺、万谊强、高嵩峰、王友军、王茵、申楠楠、时国欢。

本文件为首次发布。

1. 引言

不良资产管理领域涉及较多非标准业务，各类项目的操作方式灵活多样，各业务条线维护和关注的信息丰富繁杂。在项目进行的各个阶段和尽职调查、方案审查、风险控制、资产处置和机会发现的各个环节，业务人员需要对项目相关市场参与主体的信息及扩展关联数据有所了解和掌握，对企业经营基本状况、企业集团、投资集中度、授信额度、项目协同意愿或利益冲突、关联关系、担保风险、舆情事件等进行判断，涉及的关系网络不仅包括市场、监管、产业链、企业、干系人等外部信息，同时包括业务运转过程中形成的项目、客户、协同等内部信息，这些都构成了不良资产管理领域知识体系的一部分，而知识图谱在表示这些关联知识和基于网络关系进行分析方面具有天然的优势，并可以支持通过进一步数据挖掘赋能业务创新，因此在不良资产管理领域有着广泛的应用前景。本文件的制定，有利于指导和帮助不良资产管理领域知识图谱技术的具体实践。

不良资产管理领域知识图谱 技术要求

* 1. 范围

本文件确立了不良资产管理领域知识图谱架构，规定了相关构建技术框架及流程、技术要求、数据使用、安全要求，并界定了有关的术语、定义和缩略语。

本文件适用于不良资产管理行业使用知识图谱支持业务场景过程中涉及的数据和知识的获取、存储与处理，知识图谱管理和计算，应用系统维护相关的技术，其他领域也可参照使用。

* 1. 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。

本文件没有规范性引用文件。

* 1. 术语和定义

下列术语和定义适用于本文件。

不良资产管理领域

不良资产 non-performing assets

企业尚未处理的资产净损失和潜亏（资金）挂账，以及按财务会计制度规定应提未提资产减值准备的各类有问题资产预计损失金额。

干系人 project stakeholders

积极参与项目实施或者在项目完成后其利益可能受积极或消极影响的个人或组织。

1. 客户、用户、发起人、高层管理员、执行组织、公众或反对项目的人。

产业链 industry chain

指各个产业部门之间基于一定的技术经济关联，并依据特定的逻辑关系和时空布局关系客观形成的链条式关联关系形态。

知识图谱

知识图谱 knowledge graph

知识图谱是结构化的语义知识库，用于以符号形式描述物理世界中的概念及其相互关系，其基本组成单位是“实体-关系-实体”三元组，以及实体及其相关属性-值对，实体间通过关系相互联结，构成网状的知识结构。

知识（用于人工智能） knowledge(in artificial intelligence)

事实、事件、信念以及规则的汇集，以便于系统地使用.

[来源：GB/T 5271.28—2001 28.01.03]

对象（用于人工智能） object(in artificial intelligence)

具有一种或多种属性的物理或概念实体。

本体 ontology

在大数据语境下，约束后续各种不同层次逻辑模型的语义模型，是共享概念模型明确的形式化规范说明。

1. 本体，从本质上看，既可以是非常概括性的，也可以是极其专门化的。

[来源：GB/T 35295—2017 2.1.54]

实体 entity

机构名、地名、人名等专有名词或其他有意义的名词性短语。

关系 relationship

实体之间的语义联系。

1. 关系包括隐含语义联系。

属性 attribute

对实体的描述。

事件 event

发生在某个特定时间点或时间段、某个特定地域范围内，由一个或者多个角色参与的一个或者多个动作组成的事情或者状态的改变。

节点 node

在RDF协议框架下，指图谱中三元组的主语或宾语，个别情况下也可是谓词IRI。

标签 label

节点的分类结果信息。

三元组 triple

表示RDF中的一组关系，由主语（subject）、谓语（predicate）和宾语（object）三个部分组成。

实体链接 entity linking

为文本中提到的实体（如著名的个人、地点或公司）分配独特身份的任务。

知识抽取 knowledge acquisition

查找、收集和精化知识，并将它转换成一种形式的过程，该形式能被基于知识的系统进一步处理。

[来源：GB/T 5271.28—2001 28.01.09]

知识表示 knowledge representation

将知识编码并存入知识库的过程或结果。

[来源：GB/T 5271.28—2001 28.01.08]

知识融合 knowledge fusion

知识组织与信息融合的交叉学科，获取隐含的或有价值的新知识，优化知识的结构和内涵，提供知识服务。

1. 面向需求和创新，通过对众多分散、异构资源上知识的获取、匹配、集成、挖掘等实现。

知识推理 knowledge reasoning

按照某种策略，根据已有知识推出新知识的过程。

* 1. 缩略语

下列缩略语适用于本文件。

AMC：资产管理公司（Asset Management Companies）

IRI：国际化资源标识符（Internationalized Resource Identifier）

OWL：网络本体语言（Web Ontology Language）

RDF：资源描述框架（Resource Description Framework）

NER：命名实体识别（Named Entity Recognition）

* 1. 概况
		1. 领域相关特点

不良资产管理领域使用知识图谱，相关的领域特性有：

1. 涉及的数据种类较为综合和广泛；
2. 较为依赖行业或企业内部特有的非公开信息；
3. 业务运转天然形成网状信息，易于用图谱表示，同时应用也更依赖图谱提供的能力。
	* 1. 领域相关限制

不良资产管理领域使用知识图谱，相关的领域限制有：

1. 不良资产管理行业信息科技化以及基于人工智能和大数据的知识图谱技术应用仍在发展阶段，为保持技术路线稳健，更关注经过较为广泛使用验证和较为成熟的技术和方式；
2. 不良资产管理属于较为传统的金融行业，相对于底层技术，更关注知识图谱的业务应用；
3. 金融数据具有歧义多、噪声大、碎片化的特点，知识图谱所能吸收的高质量数据较为有限，知识图谱中的知识抽取、知识融合及知识计算等技术面临较多困难与挑战；
4. 不良资产管理领域知识图谱包含多方数据，对数据的加工和使用涉及到相关的金融安全考量。
	1. 知识图谱技术架构

不良资产管理领域知识图谱技术架构见图1：



1. 不良资产管理领域知识图谱技术架构

不良资产管理领域知识图谱技术框架由源数据、数据处理、知识抽取及知识融合、图谱中心、知识推理、应用技术及业务场景等六个层次组成。知识表示及储存贯穿整个知识图谱技术架构，负责将前三个层次获取的知识按照统一规则保存在适当媒介内，并向后三个层次提供便利快速的知识使用、知识内容更新和知识结构重组。

图谱的源数据由结构化数据、半结构化数据和非结构化数据组成，包括但不限于市场和企业基础数据和关联关系和属性数据，行业领域特色数据，内部业务相关数据等。基于数据处理层的数据载入、数据抽取、数据转化技术，为数据向知识的转换做好准备。

完成基本处理的数据，通过知识抽取相关的技术如NER、关系抽取、事件抽取、属性抽取等，结合知识融合相关的方法如本体匹配、实体对齐等，形成包括企业图谱、干系人图谱、舆情事件图谱、领域业务图谱和监管和法规图谱等在内的面向应用的图谱数据，汇集在图谱中心。

基于图谱中心数据进行搜索便利、路径发现、社区聚类、子图挖掘、实体链接和标签传播等知识推理后，为探索查询、关联发现、知识问答等应用技术提供支撑，并应用于知识可视化、穿透查询、智能客服、限额控制、风险预警、客户画像和机会发现等业务场景。

* 1. 知识图谱构建和应用
		1. 源数据

源数据和数据获取符合以下要求：

1. 可利用现有通用知识图谱、领域知识图谱中的已有知识；
2. 可从稳定可靠的外部公开合法数据源或第三方合法数据源获取权威信息，包括但不限于企业工商数据、司法诉讼数据等；
3. 可从网络公开渠道获取用于进一步分析提取信息的数据，包括但不限于舆情新闻、行业词库等；
4. 可结合内部业务自有数据、领域专家知识和外部数据构造融合数据集合；
5. 应优先使用定义清晰的结构化数据，以保证图谱知识准确可靠；
6. 对于分析处理获得的非确定性参考信息，应与确定性信息进行区分。
	* 1. 数据处理

数据处理符合以下要求：

1. 应对存在噪声的数据进行清洗操作；
2. 应按照数据存储方式的选择将原始数据转化为相应的格式；
3. 可保留用于直接导入存储的中间数据；
4. 可保留中间数据的多个版本，或对中间数据的增量和差异进行记录；
5. 对于涉及到敏感信息的数据，应符合第9章的安全要求。
	* 1. 知识抽取

知识抽取符合以下要求：

1. 可使用直接映射的方法进行结构化数据的知识抽取；
2. 可使用模板解析的方法进行半结构化数据的知识抽取；
3. 可主要使用基于领域专家构建的规则进行非结构化数据的知识抽取；
4. 可使用实体抽取算法、关系抽取算法、属性抽取算法和事件抽取算法进行非结构化数据的知识抽取；
5. 可保留用于知识抽取的映射、模板、规则以及人工智能模型的参数。
	* 1. 知识表示

知识表示符合以下要求进行：

1. 应对领域概念进行划分，对实体、关系、属性、事件等知识进行定义和格式化表示；
2. 可依据行业惯例进行知识图谱元素定义和逻辑结构构建；
3. 应根据知识图谱的数据规模、操作复杂度、模型结构来选择知识表示的方式；
4. 对于基于语义网进行的知识表示，应遵循万维网联盟（W3C）发布的各项标准，使用唯一的IRI表示资源，使用RDF框架对IRI进行描述，使用OWL本体语言对本体进行描述；
5. 对于基于特征向量进行的知识表示，可使用平移距离模型、语义匹配模型或其他机器学习模型；
6. 可使用三元组作为知识图谱表示和数据交换的基本形式。
	* 1. 知识融合

知识融合符合以下要求：

1. 应对知识图谱的概念层和数据层进行融合；
2. 可使用基于术语的方法、基于结构的方法、基于本体中的实例的方法或三者综合运用的方法进行概念层的本体匹配；
3. 可使用基于规则的方法、基于聚类的方法进行数据层的实体对齐；
4. 内外部数据的融合，涉及通用类型的数据时，可使用统一标识进行匹配；
5. 可构建不良资产管理行业术语、缩写和别名对照、词汇库等用于辅助知识融合；
6. 构建知识图谱可使用自顶向下或自底向上的方法：
	1. 使用自顶向下的方法构建知识图谱时，应先确定知识图谱的数据模型，再根据模型填充具体数据，最终形成知识图谱；
	2. 使用自底向上的方法构建知识图谱时，应首先对实体进行归纳组织，形成底层概念，然后逐步往上抽象，形成上层的概念；
7. 知识概念的更新可通过领域专家人工审核。
	* 1. 知识存储

知识存储符合以下要求进行：

1. 不良资产管理领域的实体应定义和存储为节点，例如企业、法人等；
2. 不良资产管理领域实体间的关系应定义和存储为节点间的边，例如从属、投资等；
3. 可根据知识图谱的数据规模、操作复杂度、模型结构来选择知识存储的方式；
4. 可优先使用图数据库用于知识图谱的存储，可使用关系型数据库作为辅助存储方式；
5. 数据存储可支持多副本，保证安全、可容灾、高可用等性能要求；
6. 可通过提高服务器硬件配置或服务器数量的方式来实现集群数据处理能力的提升；
7. 可支持高效自动的知识图谱数据更新策略。
	* 1. 图谱中心
			1. 图谱管理

图谱管理符合以下要求：

1. 应提供集中的方式管理图谱；
2. 可支持图谱模型的统一管理；
3. 可支持对图谱整体和子图的查看；
4. 可支持图谱的导入和导出，导入导出支持通用的标准格式；
5. 可支持对图谱中的实体、关系、属性进行查看和管理等操作；
6. 可支持图谱查询语句的统一管理；
7. 可提供可视化的方式管理图谱中心。
	* + 1. 领域图谱

领域图谱符合以下要求：

1. 应以有效支撑业务的分析和决策需求为构建目标；
2. 企业和干系人图谱可包含不良资产管理领域业务参与企业、法人、自然人等实体及其属性；
3. 企业和干系人图谱可支持实体之间的投资关系、组织关系、司法诉讼关系、经营合作关系、基于其他共同属性的关系等关系及其属性；
4. 领域业务图谱可包含不良资产管理领域业务参与方、业务标的、业务领域术语和规程等实体及其属性；
5. 领域业务图谱可支持业务参与方之间的组织关系、法律关系、经营关系等关系及其属性；
6. 领域业务图谱可支持业务标的之间的分解、组合、迁移、转化等关系及其属性；
7. 监管和法规图谱可包含不良资产管理领域业务参与实体、监管机构涉及实体、法律法规涉及实体等实体及其属性；
8. 监管和法规图谱可支持各实体之间与监管和法律法规相关事件关联的关系及其属性；
9. 舆情事件图谱可包含不良资产管理领域广泛的业务参与实体及其属性；
10. 舆情事件图谱可支持各实体之间与动态事件关联的关系及其属性；
11. 应支持领域图谱的扩展。
	* + 1. 图谱可视化

图谱可视化符合以下要求：

1. 应以直观方式提供图谱的可视化；
2. 应对实体、关系、属性进行区分；
3. 以传统的节点和连线方式表示图谱时，应以节点表示实体，以连线表示关系；
4. 应体现属性与对应实体或关系的关联；
5. 可对实体的属性和关系的属性进行区分；
6. 可对图谱中的实体进行钻取显示；
7. 可支持对特定实体间的特定关系进行显示；
8. 可支持用实体的属性对可视化内容进行筛选；
9. 可支持用关系的属性对可视化内容进行筛选；
10. 可支持用关系的层数对可视化内容进行筛选；
11. 在图谱元素数量较多时，可支持对可视化方式进行优化；
12. 可支持不同图谱之间的关联。
	* 1. 知识推理

知识推理符合以下要求：

1. 可基于不同的关系构造相关子图，用于对特定关系或事件进行查询；
2. 可支持对图谱的搜索；
3. 可支持对图谱的遍历；
4. 可支持对图谱的路径发现计算；
5. 可支持对图谱的聚类计算；
6. 可支持基于规则、统计、机器学习的知识推理。
	* 1. 知识图谱应用

知识图谱应用符合以下要求：

1. 应支持对实体、关系和属性的查询；
2. 应支持对两个或多个实体间关系、关系属性的查询；
3. 可支持对基于特定关系的图谱的查询；
4. 提供的图谱信息应包括所有必要的实体和关系；
5. 对于包含多层级关系的图谱，可支持关系的逐级查询和节点的扩展功能；
6. 可提供可视化的结果展示，并支持必要的操作；
7. 可提供图谱导出和保存能力；
8. 应具备应用接口，提供数据服务能力；
9. 系统响应时间应在人类可感知接受的范围内；
10. 根据用户权限对可见图谱范围和应用返回结果进行控制，涉及敏感数据时，应符合第9章安全要求。
	1. 知识图谱系统运维

知识图谱系统运维应符合以下要求：

1. 图谱相关的基础系统部署于企业可控的运行环境内；
2. 提供系统运行环境的管理机制；
3. 编制和维护系统运维的操作手册；
4. 编制图谱数据备份、恢复操作手册；
5. 编制系统应急响应处理手册；
6. 提供必要的手动维护功能。
	1. 安全

知识图谱安全符合以下要求：

1. 应制订图谱数据的全生命周期安全策略，包括数据采集、传输、分析、存储和销毁等阶段；
2. 数据应存储于企业可控的运行环境；
3. 可对内部和外部数据采用不同的安全策略；
4. 应对敏感数据进行脱敏；
5. 图谱软件系统应部署于企业可控的运行环境和位置；
6. 应具备权限管理功能；
7. 可接入企业认证系统；
8. 应遵循企业对安全的其他要求；
9. 可根据等保2.0标准对安全进行管控。

参考文献

[1] GB/T 5271.28—2001 信息技术 词汇 第28部分：人工智能基本概念与专家系统

[2] GB/T 35295—2017 信息技术 大数据 术语

[3] 中国电子技术标准化研究院.知识图谱标准化白皮书[R/OL].(2019-09-11) [2020-11-26]. http://www.cesi.cn/201909/5588.html

