



中华人民共和国国家标准

GB/T XXXXX—2024

机器可读标准能力等级模型

Machine readable standards capability classification model

（征求意见稿）

（本草案完成时间：2024年2月）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前 言	II
1 范围	3
2 规范性引用文件	3
3 术语、定义和缩略语	3
3.1 术语和定义	3
3.2 缩略语	3
4 机器可读标准能力模型	4
4.1 模型构成	4
4.2 机器可读标准维度	4
5 机器可读标准能力等级	4
6 机器可读标准能力要求	5
6.1 概述	5
6.2 内容创建	6
6.3 内容管理	6
6.4 内容交付	8
6.5 内容应用	10
参 考 文 献	14

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。

本文件由中国机械工业联合会提出。

本文件由全国工业过程测量控制和自动化标准化技术委员会（SAC/TC124）归口。

本文件起草单位：机械工业仪器仪表综合技术经济研究所等

本文件主要起草人：

机器可读标准能力等级模型

1 范围

本文件规定了按照机器可读能力对标准进行分级的通用模型，包括机器可读标准能力模型、机器可读标准能力等级和机器可读标准能力要求。

本文件适用于机器可读标准相关方开展标准机器可读能力的水平识别和能力提升等活动，并指导相关机构对标准机器可读的能力水平进行评估。

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 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

3 术语、定义和缩略语

下列术语和定义适用于本文件。

3.1 术语和定义

3.1.1

机器可读标准 machine-readable standard

以用户/业务为需求由机器、软件或自动化系统转化或生成的规范性文件或内容。

3.1.2

机器可读能力等级 machine-readable capability classification

以标准承载的规则、指南或特性能够通过机器进行读取、传输与使用的程度。

3.1.3

标签集 tag set

用于标识的标记集合。

3.1.4

标准信息模型 standard information model

一组具有相同含义的规范化数据对象类型及其依赖性关系。

注：信息模型也可称为语义数据模型

3.1.5

标准管理壳 standard administration shell

用模型化的方法提出标准需求、描述并交互标准内容的工具。

3.2 缩略语

下列缩略语适用于本文件。

CDD: 公共数据字典 (Common Data Dictionary)

HTML: 超文本标记语言 (HyperText Markup Language)

JSON: 对象表示法 (JavaScript Object Notation)

XML: 可扩展标记语言 (Extensible Markup Language)

4 机器可读标准能力模型

4.1 模型构成

本模型由能力维度、能力项和能力子项构成，具体模型构成如图1所示。

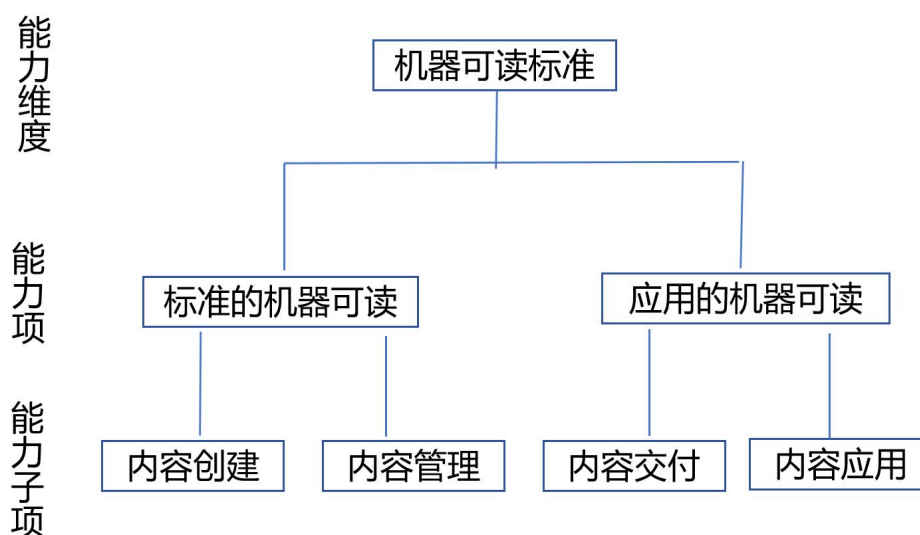


图1 模型构成

4.2 机器可读标准维度

机器可读标准维度用于描述标准化过程的全生命周期，包括内容创建、内容管理、内容交付和内容应用四个阶段：

——机器可读标准的内容创建包括两种方式，一种是借助软件工具直接将标准编写为数字化的机器可读标准，另一种是对一般形式标准进行转化，使其转化为机器可读标准；

——机器可读标准的内容管理，是对由标准信息单元及其关联关系组成的标准资源库进行管理；

——机器可读标准的内容交付，是基于标准应用的场景，提供交付物及交付平台，并保障交付质量；

——机器可读标准的内容应用，为用户提供机器可读标准的应用服务。

5 机器可读标准能力等级

机器可读标准能力等级规定了标准在机器可读的不同阶段应达到的水平。机器可读能力等级分为五个等级，自低向高分别为一级（机器可显示）、二级（机器可处理）、三级（机器可执行）、四级（机器可解析）和五级（机器可决策），如图2所示。

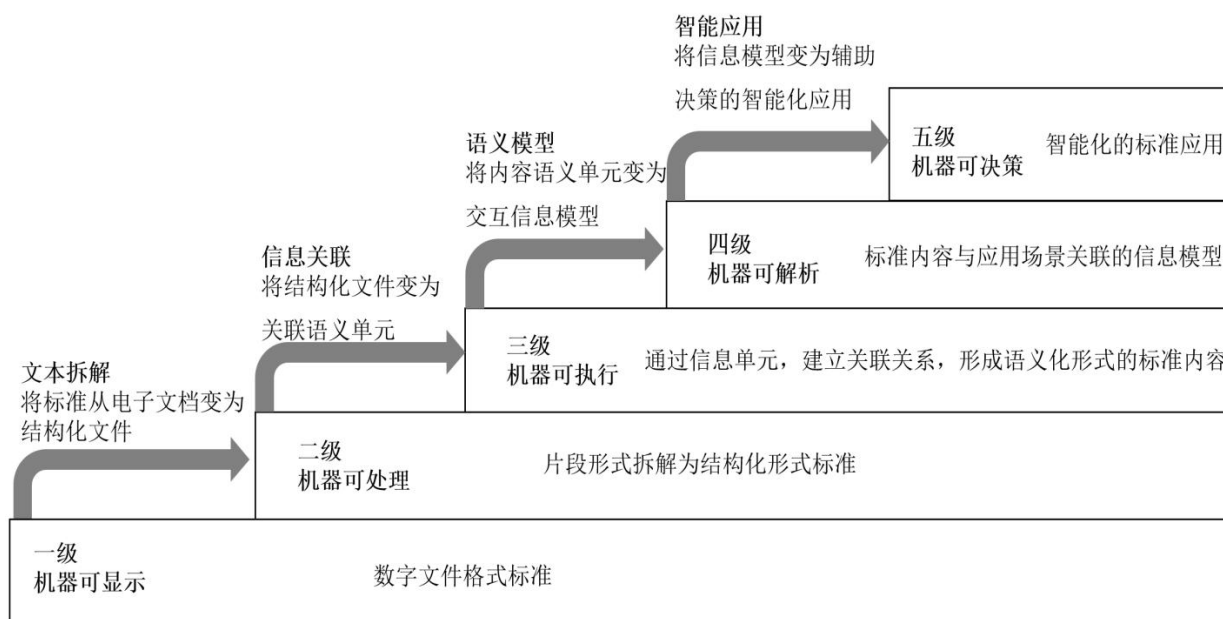


图2 机器可读标准能力等级

机器可读标准能力等级中第一级至第二级是将标准电子文档进行文本拆解为机器可处理的结构文件；第二级至第三级是将机器可处理的结构化文件转变为信息关联的语义单元，第三级至第四级是将机器可执行的语义单元转变为可交互的语义信息模型；第四级至第五级是将机器可解析的信息模型转变为辅助机器可决策的智能化应用。其中，第三级至第五级是机器可读标准的高阶能力，基于应用需求将标准内容进行加工，便于机器使用及辅助决策。每一级对应的能力要求如表1所示。

表 1 机器可读标准能力要求

能力等级	形式	机器可读能力要求
一级	机器可显示文档	机器可显示和搜索相关内容。
二级	机器可处理文件	包含标准文本结构化的内容，可利用软件识别文件结构并进行基本处理。
三级	机器可执行内容	可根据应用场景选择性地访问赋有语义的标准内容，可利用应用程序界面对标准内容执行较复杂的操作。
四级	机器可解析内容	机器能以更为复杂的方式执行或解析标准内容。包含表示标准内容及元素之间相互关系的信息模型（标准管理壳），可实现无断点、无歧义的数据流。初步具备可自主应答解析询问反馈结果的能力。
五级	机器可决策内容	由机器自动修改标准内容与自主决策采用相关标准内容，标准能够自主更新、优化适应最新技术和监管框架等约束条件。具有自学习和验证的能力，可自主应答询问、预测性地提供所需标准内容（如响应产业链采购、销售环节的需求等）。

6 机器可读标准能力要求

6.1 概述

机器可读标准能力要求规定了机器可读标准在不同等级下应满足的具体条件。

6.2 内容创建

内容创建能力子项包括需求分析、内容生成、内容验证 3 个能力特征。内容创建的等级要求见表 2。

表 2 内容创建的机器可读标准能力等级要求

能力子项	能力特征	一级	二级	三级	四级	五级
内容创建	需求分析	a. 内容调研，应通过问卷、访谈形成立项建议； b. 内容搜集，应通过现有的标准及业务需求分析标准化对象，查找资料形成电子化的比对分析文档。	a. 运用信息化手段进行标准内容调研； b. 运用信息化手段进行标准及业务需求分析，开展标准比对； c. 应通过验证平台，明确标准适用范围，确定是否为增量标准还是存量转化。	a. 运用信息化平台进行标准内容调研及大数据分析，形成立项建议； b. 运用信息化平台及大数据分析方法进行标准比对，给出比对结果； c. 应通过信息化平台及大数据分析明确标准适用范围，确定是否为增量标准还是存量转化。	a. 应集成数据分析的自动识别并梳理标准立项建议； b. 应集成智能化模型分析方法，根据标准研制周期给出比对结果及计划； c. 应集成智能化模型分析方法，形成符合业务应用需求可交互使用的标准内容。	a. 应用智能化手段根据内外部知识库信息生成立项建议； b. 应用智能化手段根据标准研制周期给出比对结果及计划； c. 应用智能化手段自动生成符合业务应用需求的标准全文。
	内容生成	应通过网络化手段生成标准化电子文档。	a. 应通过软件进行标准起草，生成 XML 等格式文件； b. 应通过协同化软件工具进行标准的起草。	应建立存量标准知识库，具备标准内容的关联引用管理信息模型，辅助标准内容生成。	应建立标准内容创建协同平台，实现标准相关利益方、管理机构等多方信息交互、协同。	应基于专家规则及用户画像，通过智能化手段，根据业务应用精准实现动态的自动生成标准内容。
	内容验证	应制定相关内容验证的规范。	应具备主要性能验证能力；基于计算机对标准内容进行分析。	应建立验证数据分析知识库，对标准内容进行验证分析。	应搭建完整的标准虚拟仿真和试验验证平台，明确标准适用范围，实现对标准技术内容的综合验证，优化标准结构。	应用多种标准验证平台协同实现基于专业的标准内容的测试验证与仿真优化。

6.3 内容管理

内容管理能力子项包括内容采集、内容加工、内容分析、内容存储、内容表达和内容传播 6 个能力

特征。内容管理的等级要求见表3。

表3 内容管理的机器可读标准能力等级要求

能力子项	能力特征	一级	二级	三级	四级	五级
内容管理	内容采集	人工采集文档，应将纸质标准输入或扫描形成PDF格式文档。	a. 应对标准内容进行结构化采集。 b. 标准内容采集范围主要在单一领域或类型。 c. 标准采集的过程需人工处理。	标准内容采集范围应涉及多个行业和类型。	a. 应对非结构化标准内容采集。 b. 标准采集范围应基本覆盖所有标准。 c. 应利用机器实时获取标准最新状态与标准最新内容。	应从机器中自动获取相关标准内容。
	内容加工	应包含标准封面、前言、引言中的关键基础信息标签（如标准号、标准名称、标准状态、发布日期、发布单位等）。	a. 应借助标记语言（HTML）对标准数据和信息进行结构化加工和标记，以便机器进行识别处理。 b. 应包含 GB/T 1.1-2020 中规定的除核心技术要素外的必备要素的基本信息（封面、目次、前言、引言、范围、规范性引用文件、术语和定义、符号和缩略语、分类和编码/系统构成、总体原则和/或总体要求）。	a. 应借助文档数据格式转换语言（XML、JSON）实现标准内容的语义化处理，形成标准资源库。 b. 应包含 GB/T 1.1-2020 中规定的必备语义要素的段落的结构化。	a. 应将标准中的结构化数据、半结构化数据和非结构化数据通过自动化技术手段提取出知识要素（实体—属性—关系），从而形成多模态数据模型。 b. 应拆解标准最小信息单元，语义层的关联关系。	应加工成一种能够表示标准内容和元素之间相互关系的格式文件，并能结合场景应用进行分析处理。
	内容分析	/	应通过软件识别加工后的文本结构。	应直接解析标准内容，分析标准信息并能够进行归类整理。	应通过知识网络实现标准知识的关联、分析和推理操作。	应在不同系统和平台之间进行标准的互操作，使标准数据能够轻易共享和交换，形成更加开放和灵活的数据生态系统。
	内容存储	应通过人工方式进行存储。	应通过含有结构化的文本文件进行存储，通过编码规则能够转换成机器可	应将标准各要素组织成树状结构进行存储，或转换成易于机器处理的数据	应通过 RDF（资源描述框架）规范存储格式进行存储（Jena），或通过	数据存储方式应采用分布式存储和计算，以适应海量数据的存储和处理需求。

			处理的文件，采用本地存储方式。	结构，采用本地存储方式。	图数据库存储（Neo4j），采用服务器存储方式，并具备数据备份和恢复能力。	
	内容表达	a. 应实现标准全文在线浏览。 b. 离散性的标准文件。	应实现标准内容结构化、可视化。	应将碎片化的标准内容按照关联关系进行展示。	a. 应通过知识地图进行智能化导航。 b. 应实现低代码、模块化标准表达。	应自动实现标准分支、标准系列、标准领域、具体标准等多维度、多视角、多模式的可视化分析与展示。
	内容传播	a. 应将标准文档存放在计算机系统中，实现对标准名称和编号的简单检索。 b. 应采用 Web1.0 的系统框架搭建标准内容传播平台。	a. 应实现对标准关键词检索。 b. 应采用 Web2.0 的系统框架搭建标准内容传播平台。	a. 应实现不同标准资源库之间的内容传输。 b. 应采用 P2P 的系统框架搭建标准内容传播平台。	a. 应实现对标准与其它应用和设备的有效集成。 b. 应利用大数据和云计算的方式搭建标准内容传播平台。	应根据用户的个性化需求提供自动化服务。

6.4 内容交付

内容交付能力子项包括交付物、交付平台和交付质量 3 个能力特征。内容交付的等级要求见表 4。

表 4 内容交付的机器可读标准能力等级要求

能力子项	能力特征	一级	二级	三级	四级	五级
内容交付	交付物	应提供符合 GB/T 1.1-2020 的电子化文档。	a. 应提供包含 GB/T 1.1-2020 要求的结构化文档，并按照结构化元模型进行拆解(如 XML 文件)； b. 应具有符合结构化元模型的标准标签集。	a. 应提供内容语义化的标准数据内容； b. 应提供标签集、拓展标签集、CDD 等； c. 应提供标准创建的过程信息。	应提供应用场景的相关数据库或数据接口。	/
	交付平台	/	a. 应具有开放的标准接口和成熟的搜索引擎；	a. 应具备语义检索能力； b. 应具备按需求进	a. 应具备根据用户人物画像和应用需求进行内容定向推	a. 应提供包含可持续发展、社会本位、预期收益等全方位

		<p>b. 应具有在线标准结构化阅读、下载功能；</p> <p>c. 应具备与文档、标准信息进行管理的功能。</p>	<p>行选择性访问的能力；</p> <p>c. 应具备信息交互关系的能力（如标准知识图谱等）；</p> <p>d. 应提供第三方系统访问标准数据的能力。</p>	<p>送的能力；</p> <p>b. 应具备识别并分析用户语义的能力</p> <p>应具备符合性检验、智能问答的能力；</p> <p>c. 应具备标准管理或信息模型，实现兼容主流设计、生产、维护、项目管理等软件及平台能力，以及接入不同异构系统的能力。</p>	<p>协同产物；</p> <p>b. 应具备人工智能分析及预测的能力。</p>
交付质量	<p>应提供符合 GB/T 1.1-2020 的电子文档。</p>	<p>a. 应保证数据格式的兼容性；</p> <p>b. 应保证结构化标准文档与原始标准的内容一致，数据传输和存储过程中数据的准确性；</p> <p>c. 应保证包括模型数据的完整性、属性数据的完整性和关联数据的完整性</p> <p>应保证数据的安全性，防止数据泄露和篡改，在数据传输和存储过程中，采取相应的安全措施（如加密、身份验证等手段）。</p>	<p>a. 应保证信息交互关系的准确性；</p> <p>b. 应保证不同来源标准数据的语义标签的一致性；</p> <p>c. 应保证过程文件的可追溯性；</p> <p>d. 应具备标准管理壳等相关信息模型库类别，标准管理壳层次清晰且完整，支持拓展。</p>	<p>a. 应保证数据信息的颗粒度与应用场景需求颗粒度的一致性（匹配度）；</p> <p>b. 应保证用户信息的安全性</p> <p>与保密性。</p>	<p>应具备运用人工智能手段，自动完成对交付内容完整性、准确性、一致性、安全性、兼容性、可追溯性等检验及优化的能力。</p>

6.5 内容应用

内容应用能力子项包括标准检索、内容推送、知识问答、内容校核、实施监督、效果评价、标准引用、标准决策支持、标准阅读、标准变更、标准匹配、标准与系统集成 12 个能力特征。内容应用的等级要求见表 5。

表 5 内容应用的机器可读标准能力等级要求

能力子项	能力特征	一级	二级	三级	四级	五级
内容应用	标准检索	应支持标准文件中的题录信息检索。	应支持对标准内容信息检索，包括但不限于标准题录信息、技术内容、图表、公式等。	应支持对标准技术内容语义检索，能够显示检索内容与检索词相似度等级；能够通过检索标准内容，显示标准关联信息；支持标准阅读过程中的知识检索。	应能通过语义识别实现描述性检索、问答式检索，内容能够精确的、细颗粒度被找到，检索过程能够不断获得机器反馈。	应支持多模态检索，检索内容多样化显示。
	内容推送	应支持标准号、标准名称的推送。	a. 应支持通过信息化系统推送标准全文； b. 应支持通过信息化系统推送标准技术内容。	a. 应支持通过信息化系统推送标准全文及标准技术内容； b. 应提供自动识别需求匹配推送标准的功能。	a. 应支持通过信息化系统推送标准全文及标准技术内容； b. 应提供自动识别需求匹配推送的需求相关标准及技术内容的功能。	应提供自动识别需求匹配推送机器可执行的需求相关技术内容的功能。
	知识问答	/	对标准的起草人、起草单位、段落、术语等进行问答。	对标准的指标、条款等技术要求内容进行问答。	a. 支持对问答场景的理解； b. 支持对特定场景的技术要求进行问答。	a. 应支持对问答的语义环境、业务场景、问答意图进行深度理解； b. 支持机器交互式多轮问答； c. 在问答中提供建议。

	内容校核	/	/	应按标准结构内容提供标准的技术要求,供人工进行校核。	a. 机器辅助提供符合性校核能力; b. 按业务场景、按标准化对象、按校核指标,对结果与标准的依从性进行校核; c. 根据校核结果提供辅助建议。	a. 机器自动判断校核场景及校核对象; b. 机器智能提供标准依从性校核结果。
	实施监督	应能分类查询标准文件的类型按需进行监督指导。	新产品、改进产品和服务、进行技术改造等,应符合标准要求并与标准文档中的内容条款实现匹配对应。	a. 具备机器可读标准的相关机器、软件和系统的可用性监督; b. 具备机器可读标准的相关机器、软件和系统的适用监督; c. 具备机器可读标准的相关机器、软件和系统的安全性监督。	a. 具备标准信息模型与业务之间的匹配; b. 具备标准管理壳组件对场景解析实现实施监督	a. 具备标准实施问题的自动决策; b. 具备标准实施和废止的自动追踪; c. 具备标准合规性的自动监督; d. 具备标准先进性的决策优化。
	效果评价	a. 具备对标准文件本身的科学性、实用性和先进性进行评价; b. 规范性、合法性、适用性、可操作性和先进性的电子化评价。	a. 具备对机器可读标准架构适用性、完整性、机器可读性的评价; b. 具备对机器可读标准内容完整性、机器可读性的评价。	a. 具备标准对于应用场景适用性的数字化评价; b. 具备标准对于应用场景的可操作性的数字化评价; c. 具备标准对于应用场景的先进性的数字化评价; d. 具备标准对于应用场景的合法性的数字化评价。	具备标准内容及元素之间相互关系的信息模型。	a. 具备标准以及标准化经济效益的数字化评估; b. 具备标准以及标准化社会效益的数字化评估; c. 具备标准化辅助建议的智能反馈。

标准引用	支持通过机器阅读查看标准规范性引用文件和内容的引用文献。	a 应支持引用标准和引用内容的动态链接； b 实现引用标准和引用内容的实时访问。	支持以语句级别的内容引用。	a. 应支持语义级内容引用； b. 应实现与具体要求相似标准内容的关联、显示和查看。	应支持数据要求的自动关联和动态链接。
标准决策支持	应为用户决策问题提供标准所属领域、要求范围和数量分布等信息。	应为用户决策问题提供相关标准文本段落内容。	应为用户提供与决策问题相关的信息。	应根据决策问题要点，提供与决策相关的标准中具体要求。	应自动识别决策问题要点，并基于决策要点给出符合标准要求的决策建议。
标准阅读	标准全文应能电子化阅读。	标准内容应能碎片化阅读。	标准应能支持交互式阅读，阅读过程可以查阅编制依据或关联标准，可通过DOI直接链接到相关标准文本。	标准内容应按需嵌入各应用场景、工具直接阅读。	应通过机器自研判，将标准内容以多模式的形式精确推送给用户。
标准变更	具备由人获取、处理标准变更信息的能力。	以不同的方式（粒度）向不同的用户组提供有关修改的信息，用户能够在线、从单一权威数据源获取标准变更信息。	标准变更信息直接呈现在在线阅读标准文档过程中。	标准变更信息直接呈现在在线阅读标准文档过程中，并附数字链接表示变更内容，变更内容可溯源且附有标准变更原因。	标准发生变更后，依托不间断的数据流使标准全链内容自动生成修改并发出预警告内容，变更内容可溯源。
标准匹配	具备由人获取、理解标准内容，人工匹配产品信息与标准内容的关联关系。	具有将标准与产品信息通过产品描述的方法（如CDD）进行匹配的能力。	具备与产品信息匹配的能力，标准和产品的匹配引入机器可读标记的机制（附加专有技术注释）。	机器应自动识别标准内容信息与产品信息进行关联，自动获取通用或领域个性化标签。	机器应自动研判产品信息与标准内容之间的匹配程度，为标准需求及缺项分析等提供决策支持。

	标准与系统集成	通过人的阅读、理解，将标准内容与应用场景按需关联、集成。	标准以数据库的形式被系统调用，主要用于标准题录信息检索，标准引用参考等。	标准内容以知识要素的形式呈现，实现标准系统和用户系统之间的固定、连续链接，集成基于标准化的API（接口）和服务，能够对其进行自动化、配置和监控。	标准以统一数据模型的形式与系统集成，模型涵盖业务、功能、通信等层级，保障标准与系统之间不间断的数据流，实现标准与系统之间的互操作性。	标准内容完全融入系统，以流程或功能固化。
--	---------	------------------------------	--------------------------------------	--	--	----------------------

参 考 文 献

- [1] GB/T 20000.1 标准化工作指南 第1部分：标准化和相关活动的通用术语
 - [2] GB/T 42093.1-2022标准文档结构化 元模型 第1部分：全文
 - [3] GB/T 42093.2-2022标准文档结构化 元模型 第2部分：技术指标
 - [4] 《FOR DIGITIZING STANDARDIZATION AND STANDARDS-WHITEPAPER》-DKE
-