

引用格式: 李景, 周洁, 李国鹏, 等. 本体技术在文献工作中的应用实践——以标准文献为例[J]. 标准科学, 2025(4):127-134.
LI Jing, ZHOU Jie, LI Guo-peng, et al. Study on Ontological Theory and Practice in Literature and Documentation - Taking Standard Documentation as an Example [J]. Standard Science, 2025 (4):127-134.

本体技术在文献工作中的应用实践 ——以标准文献为例

李景¹ 周洁¹ 李国鹏^{2*} 张明¹

(1. 中国标准化研究院; 2. 中国科学院科技战略咨询研究院)

摘要: 【目的】本体 (ontology) 技术在标准文献领域的应用和实践研究, 能够拓宽本体的应用范围, 为标准文献的知识组织提供新方法新途径, 有利于为用户提供更好的知识服务。【方法】通过文献调研法、语义分析法、文献计量学方法和系统构建的方法, 以标准文献为对象, 开展了基于本体的知识组织, 并在技术指标揭示、语料库构建和知识关联方面开展了实证研究。【结果】基于本体的知识组织为标准文献的技术指标揭示提供了符合学科规律的数据底层, 在水产品、乳制品、农产品领域的标准内容指标揭示中起到重要作用。在知识关联中, 建立了标准化对象的知识关联, 在太阳能光伏电池领域得到了有效应用。【结论】基于本体的知识组织方法应用在标准文献中有利于揭示重要技术内容、提高检索效率, 改善用户体验, 相关成果已经成为国家标准。

关键词: 本体; 标准文献; 知识组织体系 (KOS)

DOI编码: 10.3969/j.issn.1674-5698.2025.04.016

Study on Ontological Theory and Practice in Literature and Documentation - Taking Standard Documentation as an Example

LI Jing¹ ZHOU Jie¹ LI Guo-peng^{2*} ZHANG Ming¹

(1. China National Institute of Standardization;

2. Institutes of Science and Development, Chinese Academy of Sciences)

Abstract: [Objective] The application of and practical research on ontological technology in the field of standard literature can broaden the scope of ontology application, provide new methods and approaches for knowledge organization of standard literature, and provide better knowledge services for users. [Methods] Through literature research, semantic analysis, bibliometric methods, and system construction methods, ontology-based knowledge organization is carried out with standard literature as the object, and empirical research is conducted in terms of technical indicator disclosure, corpus construction, and

基金项目: 本文受中国标准化研究院基本科研业务费项目“数字标准馆标准体系构建及关键标准研制与应用”(项目编号252023Y-10411); 2021年国家科技图书文献中心 (NSTL) 专项项目“美国先进制造业标准化发展趋势研究”(项目编号2021XM59-3-101) 资助。

作者简介: 李景, 博士, 博士后, 研究馆员, 研究方向为本体技术、文献计量学、标准文献。

周洁, 本科, 研究馆员, 研究方向为标准文献。

李国鹏, 通信作者, 硕士, 副研究员, 研究方向为知识图谱、人工智能。

张明, 本科, 副研究馆员, 研究方向为标准大数据、技术性贸易措施、标准实施本体技术。

knowledge association. [Results] The ontology-based knowledge organization provides a data infrastructure that conforms to disciplinary rules for the disclosure of technical indicators in standard literature, and plays an important role in the disclosure of standard content indicators in the fields of aquatic products, dairy products, and agricultural products. In knowledge association, standardized object knowledge association has been established and effectively applied in the field of solar photovoltaic cells. [Conclusion] The application of ontology-based knowledge organization methods in standard literature is beneficial for revealing important technical content, improving retrieval efficiency, and enhancing user experience. Relevant achievements have become national standards.

Keywords: ontology, standard literature, knowledge organization system (KOS)

1 本体基本理论

本体的概念来自哲学领域。最早开始引入本体概念的是人工智能(AI)领域。本体可用来作为知识表示和知识组织的底层,可以规定概念与概念之间的复杂关联,并定义概念的属性。本体的概念和方法进入到知识组织和图书情报领域是在2000年初。引入本体的目的是要解决信息检索、资源组织等问题。

1.1 本体的概念

斯坦福大学的Gruber^[1]最早提出本体的定义,其后Borst Pim^[2]对Gruber的定义做出修正。两个定义合并以后为:“It is an explicit formal specification of a shared conceptualization.”中文含义:本体是一套得到大多数人认同的、关于概念体系的明确的、形式化的规范说明。

使用最广泛的本体定义是波音公司Michael Uschold等^[3]提出的:“It is a vocabulary of terms and some specification of their meaning.”中文含义:本体是一套术语词表以及术语含义的规范说明。

通过对国内外学者观点的归纳,笔者早在其博士论文(2004)中就指出,本体的定义虽然有诸多表述,但是有关它的必要条件也应该在定义中得到体现。笔者认为:本体是一个关于某些主题的、层次清晰的规范说明。是一个已经得到公认的形式化的知识表示体系,它包含词表(或名称表/术语表)。词表中的逻辑声明全部是用来描述这些术语的含义和术语间关系的,即它们是怎样和其他术

语相关联的。

1.2 本体的作用

本体能够作为不同系统之间进行互操作的逻辑底层,构建本体的作用在于:

(1) 在用户间或软件代理间达成对于知识组织表达的共同理解和认知^[1,4]。

(2) 可对专业术语、常识、通用知识,计量单位进行复用。例如,在农业、生物医药、中医药,环境保护等多个领域中,植物学基础知识都是通用基础必备知识。如果以植物分类法为基础框架,以植物学词表作为概念表示的基础,以此来构建植物学本体,则该本体可以在农业、生物医药、中医药和环保等多个领域中被复用。如果需要构建一个大规模本体,可以将几个已有本体进行集成,还可以复用现成的通用本体框架,如UNSPSC ontology,将UNSPSC ontology (Skeleton) 进行扩展和填充,就可以应用到其他领域中去^[5]。

(3) 通过断言(如三元组、逻辑表达式)等形式(公式),使专业领域内的学术观点和假设变得更加明确。对专业领域知识进行明确的形式化说明,对于那些必须理解该领域专业术语含义的新用户是十分必要的。

(4) 将领域知识从知识管理的环境中剥离出来。用户可按照必需的规范说明和执行程序来实现检索和查询等功能,而用户本身不必非要精通IT技术或各种算法^[6]。

(5) 分析专业领域的知识体系结构。在复用和扩展现有本体的尝试中,对术语进行规范的形

式语义分析,具有重要价值和意义^[7]。

1.3 本体的研究进展

1.3.1 国外研究进展

从20世纪90年代至今,三十多年来,欧美发达国家持续地支持基于本体的知识组织体系建设与开发。近十年来,更加重视知识组织体系在网络环境下的发展与应用,如普林斯顿大学开发了WordNet(初级本体),美国国会图书馆开发了Classification Web, OCLC开发了Dewey Web, 斯坦福大学开发了Ontolingua、Protégé本体开发工具,以及在医学领域享有盛名的MeSH和UMLS等。这些知识组织体系已成为网络环境下信息开发利用必不可少的重要基础设施和工具,而且正逐步从西文向多语言信息处理方向发展,其抢占信息处理主导地位的战略意图显而易见。

著名的基因本体由美国基因本体协会创建并维护,在生物信息学领域中得到广泛使用,业界利用GO提供的公共词汇和语义描述基因产品,在不同数据库中能够体现基因产品的一致性描述。它主要包括3个分支:生物过程、分子功能和细胞组件。GO包含词汇量90,261条^[8]。

在数字经济和全球化浪潮的席卷下,全球各大出版业巨头都被裹挟在数字出版的浪潮之中。不管是出于主动还是被动,都在通过网络化终端介入并渗透到对最终用户的服务中,这种行径不管是有意识或无意识,都已经抢占了传统图书情报机构的服务领地。几乎所有的数字出版商都在面向最终用户推出以内容为核心的智能检索、关联揭示、数据挖掘与数据分析等知识服务。我国是科技文献出版和使用的大国,国际出版界系统通过垄断最终用户提供知识服务,从而实现对我国文献信息服务市场的控制的战略意图。

1.3.2 国内研究进展

2006—2008年,国内在本体技术领域中的研究热点包括:本体工具、本体学习、本体库、本体提取、本体推理,本体翻译和本体语言等。国内的学者和机构更倾向于算法、匹配、映射、建模,推理和抽取研究。所有相关研究中,语义分析和语义

标引都是关键性工作。2008年之后,单纯以本体理论为研究主题的项目不再列入基金项目申请的范畴^[9]。

2006—2009年的科技部支撑计划项目“农村知识本体构建的共性技术研究”(2006BAD10A050104)中,发表论文基于花卉学科技文献的大规模本体开发环境(LODE)的开发,并实现了基于本体知识库的半自动标引(见图1)。

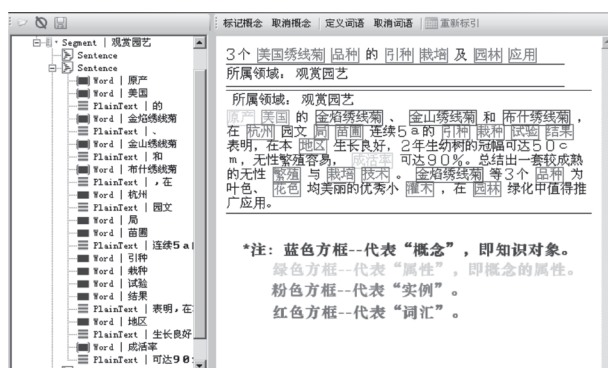


图1 对中文科技文献进行半自动语义标引的示例图
(系统自动生成页面)

2011年,王颖^[10]在中科院成都计算机应用研究所读博士期间,完成了《中医药领域本体及知识库的研究与应用》,对基于本体的中医诊断系统进行了实证研究(见图2)。

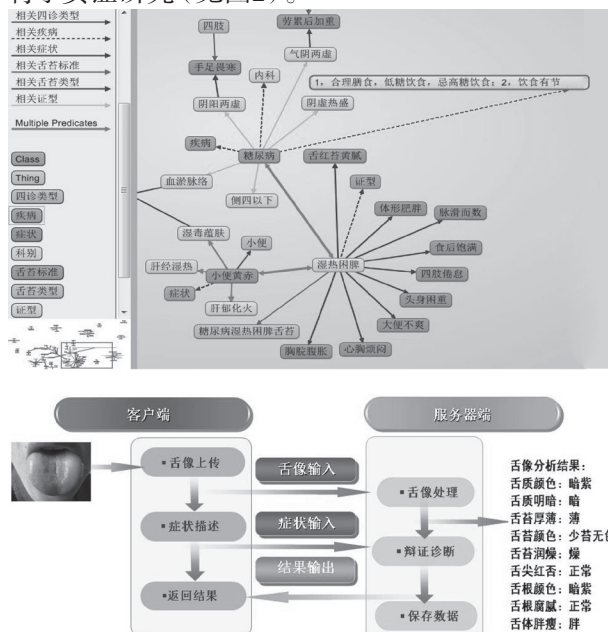


图2 中医药相关知识点的语义网结构示例和舌诊辨证流程

2011年,国家科技图书文献中心(NSTL)承担了科技部支撑计划项目“面向外文科技文献信息的知识组织体系建设与应用示范”(2011BAH10B00,又称“超级科技词表项目”)。项目的建设目标是:采用国际上先进的知识组织技术和方法,在国内外已有系统建设成果和应用经验的基础上,争取在“十二五”期间基本建成面向计算机应用的科技知识组织体系,为我国海量外文科技文献信息的组织和利用提供支撑,实现国家科技文献信息战略资源的有效组织,提供知识检索服务,揭示更深层次的知识关联,从而实现基于国家科技文献信息战略资源的知识发现、知识挖掘和知识计算应用示范,整体提升我国科技文献信息机构的知识服务能力。其具体目标中就包含:构建以领域本体为目标的外文科技知识组织体系,以及建立国家科技知识组织体系的可持续发展机制。

由此可见,2010年后,本体相关项目更侧重于构建基于本体的知识组织体系(又称“实用分类体系”),并侧重于知识组织体系的应用和实践。

2 本体理论在文献工作中发挥的作用

2.1 本体是唯一可以作为系统间互操作的底层知识库

本体中至少要包括其概念和概念间关联,这是本体的核心。如果只具备简单的概念和概念间上下位关联,如“梅花”,其上位概念为“李属植物”,“李属植物”的上位概念是“蔷薇科植物”,“梅花”的下位概念有“骨红垂枝”(梅花的品种之一)、“白碧垂枝”(梅花的品种之一),那么此类本体仅属于轻量级本体。重量级本体中包括对概念的描述,以及以形式逻辑公式表示的概念间关系。最重要的,其间包括的概念不仅仅是简单的术语表达和描述,而是多个属性值的集合。例如概念“梅花”,对概念的描述,不仅仅是蔷薇科李属植物。具体到属性,仍然有包括同义词、别名、拉丁学名,上位分类,花期,品种,花色,栽培规程,经济用途等多个属性,且每个属性具备不同的属性

值。鉴于本体中概念的关系可以反映词汇之间语义的映射关系,所以在智能检索、文献的结构化拆分、语义标注、机器翻译,以及增强大模型性能这些方面,本体都具备无可替代的作用。但是本体的生成不能完全依赖于人工智能,其底层数据和知识链接的建立,目前仍然主要依赖于传统数据加工业务中基础数据的完备和准确^[11-15]。

2.2 本体在标准文献工作中的作用

标准文献属于科技文献范畴,是一种特种文献,以其更新作废替代周期快(标准具有标龄)、技术内容含量集中、数据公式图表多、不同品种标准文本的体例格式有自己固定格式且各不相同等特点,有别于一般科技文献。基于上述特征,本体理论应用于标准文献服务工作中,能够起到以下作用:

(1) 提高用户对标准文献数据库的检索效率(查全率、查准率)。建立基于本体的知识组织体系,作为标准文献数据库的检索底层,并结合对标准文本的结构化处理,以及对标准文本所进行的知识对象(知识要素)标引,能够最大限度地揭示标准文献内容。特别是能够揭示标准文献中的技术指标,并将指标进行比对,形成比对表,能够令用户充分了解同类产品在不同品种标准中的指标基本情况,以此判定国际标准和不同国家的标准对同一产品规格要求的异同。能够令用户充分了解同类产品在同一系列标准的不同制定年代版本中的不同规定,以此掌握该产品在不同时期的行业技术发展走势。

(2) 在标准文献数据库中建立基于本体的知识关联,并基于此为用户提供组合式检索、打包服务和专题服务。

(3) 应用于参考咨询服务中(专家系统),以基于本体知识库的累积和更新,基于本体中概念间逻辑关系判断,自动为用户提供答案,并在自学习机制和自动校验机制下,不断积累、更新,使得答案不断完善,质量不断提高。

(4) 在标准文献资源建设中,基于本体的底层标准文献知识库的建立(作为提供各种应用的基础)可以为标准文献加工中的自动化概念标引提

供基础,并由此提升加工效率、拓展加工深度、改善加工准确度。基于本体的标准文献知识库还能够为多语种标准文献机器翻译提供丰富的术语翻译选择并成为多语种概念间的映射。

3 实践研究

3.1 标准文献语料库建设

语料库在语言学上指大量的文本经过整理和加工,具有规范的格式与特定标记,是存放原始语言材料的数据仓库。一般来说,语料库的语料来源可能会非常广泛,包括词典、丛书、专著、论文集等。标准文献语料库的来源主要包括标准和技术法规,以及技贸措施相关信息。知识库是用于知识管理的一种特殊的工具,用于领域知识的采集、整理和提炼揭示。知识库中的知识对象是专业领域知识的集合,包括概念、事实、规则和其他有用信息。

语料库、知识库和知识组织体系(KOS)的构建,是知识管理和知识服务的基础、前沿和热点。语料库的用途包括实现数据挖掘、语义标引、机器翻译、知识关联、智能检索等功能和模块^[16-18]。

该项研究中研发的语料库原型系统,实现了按照简单的知识组织体系(初级本体)进行专业导航的功能(见图3)。并在数据结构中设置了“知识对象”这一字段,其作用是为建立复杂的基于本体的知识库建立“知识对象”—“中文语料”—“英文语料”—“XX语种语料”奠定映射表基础。以支持后续机器翻译、自动标引等多种功能的拓展。



图3 分专业浏览语料的原型系统页面

3.2 标准文献内容指标揭示数据库建设

通过对我国食品、农产品、环保领域相关的国家标准、行业标准以及法律法规文件的收集、分类、整理、数据加工,完成了1206个食品与农产品领域、3777个环保领域标准和法律法规的内容解释、数据加工,建设了相应的数据库。与传统的以查找文本为目的的检索方法相比,除直接查询技术指标和关键技术内容外,还可获取相关系列产品和相关标准体系,提高了技术标准查询精度与效率,使检索结果有更大的利用价值。

传统检索方式,主要根据主题词、分类号等进行检索,检索结果的反馈集中体现在篇目信息(即文献的标题),不能针对标准正文内容中所涉及的具体指标和关键技术内容进行检索。以“乳制品中六六六残留量规定”这一检索目标为例:

如果使用普通的主题词检索,设定检索词为“六六六”和“乳制品”,仅得到一个结果,即《出口乳及乳制品中“六六六”“滴滴涕”残留量检验方法》(SN 0127—1992,已作废的商检行业标准)。如果输入检索词“乳制品”,得到32条乳制品国家标准与行业标准,将其中有可能涉及卫生指标的标准逐篇进行阅读分析,查找对六六六残留量的限定值。这样不但查找速度慢,而且因为无法实现系列乳制品产品标准的体系查询,检索结果可能会存在较多缺失。

如果使用标准内容揭示系统进行检索,输入检索词“乳制品”和“六六六”。因为内容揭示检索对技术指标进行了分解,而且利用本体框架对系列产品进行了分类,所以得出国家标准与行业标准给出的8个不同乳制品产品中六六六残留量的限定值,检索结果定位准确完备。如表1所示。

在构建标准文献揭示指标检索体系的过程中,吸纳了本体技术的基本方法,构建了以专用本体分类表为基础,以概念属性关系描述为重要手段的标准知识内容与技术指标的揭示检索系统,解决了标准文献内容指标的体系查询与关联检索的问题。

在研究的基础上,重点开展了食品、农产品、

表1 乳制品国内标准技术指标列表

标准种类	检索对象	属性类型	技术指标	内容	指标来源
中国-国标	乳脂肪	产品标准-卫生指标	六六六（以脂肪计）	$\leq 0.5 \text{ mg/kg}$	GB/T 5415—2008
中国-农业	纯酸乳（无公害食品）	产品标准-卫生指标	六六六	$\leq 0.01 \text{ mg/kg}$	NY 5142—2002
中国-农业	调味酸乳（无公害食品）	产品标准-卫生指标	六六六	$\leq 0.01 \text{ mg/kg}$	NY 5142—2002
中国-农业	果味酸乳（无公害食品）	产品标准-卫生指标	六六六	$\leq 0.01 \text{ mg/kg}$	NY 5142—2002
中国-国标	液态乳	卫生标准-卫生指标	六六六	$\leq 0.02 \text{ mg/kg}$	GB 19301—2003
中国-国标	稀奶油	卫生标准-卫生指标	六六六（以脂肪计）	$\leq 0.5 \text{ mg/kg}$	GB 19646—2005
中国-国标	奶油	卫生标准-卫生指标	六六六（以脂肪计）	$\leq 0.5 \text{ mg/kg}$	GB 19646—2005
中国-国标	无水奶油	卫生标准-卫生指标	六六六（以脂肪计）	$\leq 0.5 \text{ mg/kg}$	GB 19646—2005

环保领域部分重要标准内容揭示数据库建设。建立了标准文献内容揭示基础支撑数据库,包括:揭示词表数据库、本体类表数据库、揭示范畴表数据库、中标分类数据库、国际分类法数据库、标准体系表数据库。建立了应用支撑数据库,包括:主题对象数据库、标准体例数据库、计量单位数据库、限定类数据库和技术指标揭示数据库等,实现了标准中特定技术指标的检索、系列产品的检索,以及与产品相关的技术标准体系的检索。检索效果见图4、图5。



图4 主题“荔枝”指标“毒死蜱”检索界面

标准类型	检索对象	属性类型	技术指标	内容	内容来源	来源	格式标准
中国-农业	荔枝 (无公害食品)	产品标准-卫生指标	毒死蜱	$< 1 \text{ mg/kg}$		NY 2173-2002	格式标准
中国-农业	荔枝	产品标准-卫生指标	毒死蜱	$< 0.1 \text{ mg/kg}$		NY/T 215-2002	格式标准
中国-农业	荔枝 (绿色食品)	产品标准-卫生指标	毒死蜱	$< 0.5 \text{ mg/kg}$		NY/T 215-2002	格式标准

图5 荔枝(主题对象)的“毒死蜱”(指标)含量比对结果

3.3 基于初级本体的专题集成

按照已有的分类体系(如标准分类法、海关分类编码等)建立家电产品初级本体,并将相关标准文献按照已经建立的家电产品本体进行分类归纳(见图6)。按照这种方式建立的专题集成,其优势是便于查询和浏览,符合一般读者的阅读习惯。

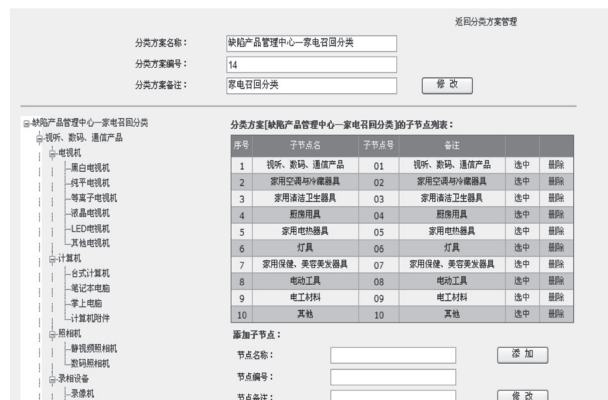


图6 家电产品专题集成产品分类页面

3.4 标准化术语质量分析与语料库应用研究

继“标准文献语料库建设(3.1)”之后,在以下方面开展探索:

(1)在语料记录数上,从现有的4500条增加到10,000条,并继续以基础标准化领域、家电和电工领域的标准文献为基础,进行语料的提取和加工。(2)现有的已建成的语料库原型系统中的语料素材,基本是以术语为主,在II期中,将尝试对自由词、词组、短句、标准文献中常见的标准句型进行基于本体中知识对象的著录标引加工。因为术语只是语料素材的一部分而已,要想进行翻译,后面

所说的这些(自由词、词组、短句、标准文献中常见的标准句型)才是更重要和有价值的,也是标准文献语料加工中没有开展过尝试的。(3)2011年开发出的语料库原型系统(prototype),只具备简单的语料记录查询功能和语料录入功能;后续将开发完备的语料记录数据库管理系统(初具雏形的基于本体的知识库),具有简单检索、高级查询和标准文献数据库的接口功能,输入和编辑复杂语料记录(不仅仅是术语)的功能,并预留和翻译平台软件的接口。(4)后续,将尝试和成熟的翻译平台软件进行对接,验证机器翻译效果。

3.5 标准技术指标与产品基础信息知识关联关键标准研究

2013—2017年依托质检行业公益专项“标准技术指标与产品基础信息知识关联关键标准研究”主要开展了基于产品基础信息的标准技术指标抽取分析统计等方法的研制,研究通用的标准技术指标体系分布规律,通用普适的分析方法和模式,建立对比规范和元数据加工规范;开展产品标准技术指标跟踪技术的研究,并建立指标跟踪模型。之后,在建立和完善标准技术指标体系与产品标准技术指标跟踪模型的基础上,以可再生能源相关产品(如太阳能光伏发电产品)企业和产品

标准技术指标体系为例,完成产品、标准、指标与企业信息等知识要素关联(基于本体的)概念模型和产品标准指标数据统计分析模型的研发和构建工作。开展了业务功能测试和试点工作,检验并测试概念模型的逻辑性和原型系统的实用性。太阳能光伏电池与技术标准和企业形成的轻量级本体逻辑关系图如图7所示。

目前,相关成果和基于标准文献的技术指标揭示技术已经成功申报,并于2021年度正式出版成为国家标准GB/T 39872—2021《标准文献技术指标揭示数据规范》^[19]。

4 结语

综合上述,在相关研究中,“标准化术语质量分析与语料库应用研究”注重基于本体的标准文献知识库的构建。“技术标准知识关联研究”注重基于本体的知识关联关键技术的研发。“标准技术指标与产品基础信息知识关联关键标准研究”注重基于本体的技术标准内容指标揭示和知识关联研究,并体现“企业-产品-标准-技术指标”这四大类知识要素之间知识关联的展示和应用。

结合企业和市场需求,以上研究会持续推进

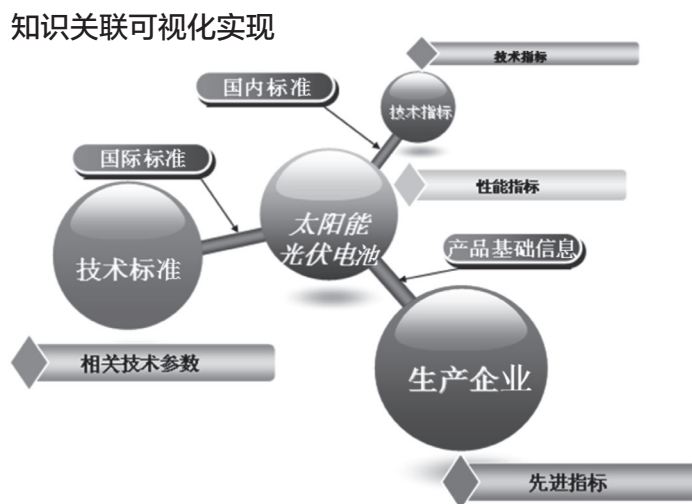


图7 太阳能光伏电池与标准和企业的本体逻辑关系图

和支持标准文献翻译、面向服务的标准文献知识关联以及标准文献知识库构建方面的工作。

基于本体的知识组织体系建设和相应的关键技术研发将会极大地促进标准文献信息系统知识服务能力的提升;促进标准文献检索从传统检索查询模式向具有内容相关揭示、跨库检索、多语言

检索等功能的智慧模式转变;促进标准文献服务工作从传统的标准文献信息提供向科技对象实体的关联发现、热点监测、趋势分析的知识服务进行转变;促进标准文献和科技文献的融合与关联,发挥合力,更好地为国家科技图书文献中心(NSTL)用户群体提供集成化、多元化的知识服务。

参考文献

- [1] GRUBER T R.A. translation approach to portable ontology specification [J]. Knowledge Acquisition, 1993(5): 199–220.
- [2] BORST P,AKKERMANS H. An ontology approach to product disassembly [P]. EKAW 1997, Sant Feliu de Gu5xols, Spain, October: 15~19.
- [3] USCHOLD M,MICHAEL G. Ontologies: Principles, methods and applications [J]. Knowledge Engineering Review. 1996 (2): 93–136.
- [4] MUSEN M A. Dimensions of knowledge sharing and reuse [J]. Computers and Biomedical Research, 1992(25): 435–467.
- [5] UNSPSC. United nations standard products and services code [EB/OL].(2003–02–01)[2025–01–10]. <http://www.unspsc.org/>.
- [6] MCGUINNESS D L,WRIGHT J. Conceptual Modeling for Configuration: A description Logic-based Approach [J]. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing – special issue on Configuration.1998(4): 333–344.
- [7] MCGUINNESS D L, FIKES R, RICE J.et al. An environment for merging and testing large ontologies. principles of knowledge representation and reasoning [P]: Proceedings of the Seventh International Conference (KR2000). A. G. Cohn, F. Giunchiglia and B. Selman, editors. San Francisco, CA, Morgan Kaufmann Publishers.
- [8] ASHBURNER M. Gene ontology: Tool for the unification of biology[J]. Nat. Genet,2000(1): 25–29.
- [9] 李景, 孟宪学, 苏晓路. 领域本体的构建方法与应用研究[M].北京:中国农业科学技术出版社,2009.
- [10] 王颖. 中医药领域本体及知识库的研究与应用[D].成都: 中国科学院成都计算机应用研究所,2011.
- [11] 张晓林. 走向知识服务:第一版 [M]. 成都: 四川大学出版社, 2001.
- [12] 李景, 钱平, 苏晓鹭. 构建基于Ontology的知识门户 [J]. 现代图书情报技术,2004 (2) : 18–21+41.
- [13] 李景. 多语种文献数据库的设计考虑 [J]. 情报学报, 2003 (3) : 321–324.
- [14] 刘晓霞. 概念图知识表示方法的研究 [J]. 计算机应用与软件2001 (8) : 56–59
- [15] 董振东, 董强. 知网和汉语研究 [J]. 当代语言学, 2001(1):33–44+77.
- [16] 教育部语言文字应用研究所计算机语言学研究室.语料库在线[EB/OL].(2010–11–10)[2025–01–12]. <http://124.207.106.21:8080/>.
- [17] 北京大学中国语言学研究研究中心.汉英双语语料库[EB/OL]. (2011–01–10)[2025–01–15].http://ccl.pku.edu.cn/yuliao_contents.asp.
- [18] 清华大学. 汉语均衡语料库.TH-ACorpus [EB/OL]. (2012–01–10)[2025–01–15]. <http://www.lits.tsinghua.edu.cn/ainlp/source.htm>.
- [19] 全国科技平台标准化技术委员会. 标准文献技术指标揭示数据规范: GB/T 39872—2021 [S].北京:中国标准出版社, 2021.