

基于MediaWiki实现知识组织系统的 语义在线维护^{*}

——以ISO 5127术语国际标准为例

黄华军 曾新红

(深圳大学图书馆, 深圳 518060)

摘要: 本文旨在为知识组织系统实现语义在线维护提供切实可行的可选方案和系统实证示范。研究了维基百科开源工具MediaWiki的特征、优势和应用, 以及ISO 5127的SKOS语义化描述整体信息和结构特征, 提出基于MediaWiki的国际化语义在线维护方案。根据ISO 5127的在线维护需求, 采用MediaWiki强大的扩展功能, 实现ISO 5127的在线维护系统, 包括术语展示、搜索和获取、用户提交建议、建议处理、实时增删改、批量导入导出、系统日志、多语言、权限管理等服务功能。本文实现的语义在线维护系统可基本满足最终用户的全方位要求, 其他类型的知识组织系统(如分类法、范畴表、规范代码表等)可借鉴此方案快速完成在线语义维护系统的实施和动态完善。

关键词: MediaWiki; 知识组织系统; ISO 5127; 语义在线维护

中图分类号: C04; TP301.2 **DOI:** 10.3772/j.issn.1673-2286.2021.10.004

引文格式: 黄华军, 曾新红. 基于MediaWiki实现知识组织系统的语义在线维护——以ISO 5127术语国际标准为例[J]. 数字图书馆论坛, 2021 (10) : 25-30.

如何实现知识组织系统的共建和及时更新维护是知识组织系统相关各界共同面对的难题, 如拥有大量高质量传统知识组织系统的图书情报界, 希望构建内部KOS以提高信息处理、知识管理和系统之间互操作效率的企业界, 以及试图构建具有强大语义处理能力与开放互联能力的知识图谱相关各界等。ISO等国际或国家标准化组织也同样面临术语等知识组织系统标准的共建和及时更新问题。

《信息与文献——基础与术语》(Information and Documentation—Foundation and Vocabulary, ISO 5127) 为整个信息领域的文献领域提供了一个概念系统和通用术语词表^[1]。ISO 5127术语国际标准新版本的修订工作由ISO/TC46/WG4负责, 修订工作的具体情况请参见文献[2]。笔者作为ISO/TC46/WG4的中方专家和ISO/TC46/AHG “Maintenance of ISO 5127” (ISO 5127

维护特别工作组) 的召集人, 负责审查可能的解决方案, 将ISO 5127作为一项数据库标准进行维护。在实现该国际标准的SKOS语义描述方案和共享服务系统ISO 5127_CLSS^[3]之后, 进一步为实现其网络在线维护和动态完善寻求可行方案。

MediaWiki最初是为维基百科量身打造的开源工具, 它自由开放、功能强大、扩展丰富。基于其在共建方面的明显优势, MediaWiki已在很多领域被广泛应用, 国内外已建设了大量Wiki项目和网站进行知识的收集和管理。

近一年多来, 笔者调研了ISO 5127国际术语标准的共建共享、在线修订和语义维护等需求, 在已有的研究成果和实践经验基础上, 借助MediaWiki的强大功能, 实现了ISO 5127的术语在线维护系统。

^{*} 本研究得到国家社会科学基金项目“中文知识组织系统形式化语义描述标准体系研究”(编号: 12BTQ045) 资助。

1 研究背景与研究意义

1.1 研究背景

知识组织系统必须不断更新才能跟上时代的发展,满足相关领域的应用需求。虽然已有少量共建研究和实施成功案例(如深圳大学图书馆NKOS研究室的OTCSS方案^[4-5]和《汉语主题词表(工程技术卷)》^[6]的叙词表协同编制软件),国内外现有的知识组织系统大多数仍单纯依赖专家维护,多由指定专家主导编辑,经过修订、审核和定稿的漫长过程,最终由出版商印刷、销售,相关工作无法实现高效的协作和去中心化,不能集中广大用户群体的意见和智慧,存在周期长、可追溯性差、获取和利用难等困扰。随着信息技术的发展,自然语言学习等机器学习方法逐渐被引入,但目前这些自动处理方法推送的结果距离真正满足高端应用需求还有不小的差距。在这种背景下,利用MediaWiki的强大共建功能快速实现知识组织系统的网络化语义在线维护不失为一种低投入、高效率的可自增值方案。

例如,ISO术语国际标准(术语表是知识组织系统的一种类型,参见参考文献[5]图1-1)传统的纸版和PDF等电子版阅读服务方式以及由少量ISO专家主导的修订工作,已严重滞后于当前网络信息时代的发展需求,这使得术语标准采用“标准作者+利益相关者”的多渠道在线维护方式越来越成为各方的共识和要求。

在ISO中引入和实现SMART (Standards Machine Applicable, Readable and Transferable) 标准是ISO未来工作的方向,其路线图(roadmap)中包括:在机器可读标准的执行标准化过程的方法上允许TC拥有自由,如工具和技术的选择,因为这种开发需要数字化的工作流程;更新法律框架以启用机器可读标准,使最终用户能够利用机器可读标准而不违反ISO法律权利;与标准利益相关者(作者和最终用户)一起制定创建机器可读标准所需的工具和过程要求,制定流程和政策,让最终用户(和外部各方)参与到机器可读标准的输入中,包括评论、更新和勘误^[7]。

因此,ISO 5127术语国际标准新版本除了应该是一种机器可用、可读和可转换的标准,即符合SMART趋势之外,还应该建立起一套有效的允许“利益相关者”参与意见的在线维护机制,以此来满足其利益相关者的诉求。

1.2 研究意义

本研究旨在借助MediaWiki的强大共建功能,为相关行业领域的知识组织系统实现语义在线维护提供实证和示范,提供一套具有实际参考和借鉴意义的解决方案,让ISO术语国际标准及其他类型的知识组织系统能够更方便快捷地实现在线维护。

此方案为用户提供广泛参与ISO术语国际标准等知识组织系统的网络化共建共享服务的便捷渠道,利用Web用户的集体智慧进一步完善这些KOS,使其走出传统的服务模式,焕发出新的生命力,从而更好地发挥其在网络时代传播知识的职能。

2 MediaWiki与ISO 5127在线维护需求的结合

本节将MediaWiki的应用特征和功能与ISO 5127的修订业务工作紧密结合,寻求在MediaWiki的基础上解决ISO 5127的在线维护难题。

2.1 MediaWiki简介

MediaWiki是一个开源的、充满活力的网络协作平台社区工具,已被成千上万的网站、公司和机构使用。它为维基百科(作为MediaWiki应用最成功的案例之一)网络协作平台提供了非常强大的动力。MediaWiki帮助收集和组织知识,并将其提供给人们使用。它具有功能强大、多语言、免费开放、可扩展、可定制、可靠且安全等特征^[8]。

国内外成功应用MediaWiki的案例,除了MediaWiki和Semantic MediaWiki^[9]本身以及维基百科之外,还有如研究数据联盟(Research Data Alliance, RDA)的TeD-T^[10]、Discourse DB(一个收集世界记者和评论员对当前政治事件和问题的意见的Wiki)^[11]、食品伙伴网旗下食品论坛运营的食品百科全书计划^[12]和专注于经济管理领域的MBA智库百科^[13]等典型例子。它们借助MediaWiki的强大能力和易用性,实现人人可自由访问和编辑、互动和贡献的开放共享社区。但目前还未见有专门面向知识组织系统语义在线维护的应用实例。

MediaWiki的应用特征和功能可通过“Help about MediaWiki”了解,其提供详细的多语种的功能介绍和用户帮助,网址详见参考文献[14]。

2.2 ISO 5127在线维护需求

国际标准化组织ISO的Online Browsing Platform (OBP)^[15]提供最新的ISO标准、图形符号、代码、术语和定义的在线搜索和访问服务,并可在标准之间轻松导航。因其内容在数据库中进行封闭维护,修订和更新在内部完成而不对外开放,无法拓宽其服务渠道和领域,网络用户也无法参与共建。

根据ISO/TC46/WG4在ISO/TC46的2019年年会中展示的ISO 5127 Annex B“ISO 5127 management of terminology maintenance (术语维护管理)”^[16]中建立的维护原则,并顺应ISO未来的SMART发展方向^[7],为ISO 5127在线维护系统设计了以下功能。(1)在线提交维护建议:ISO/TC46的标准制定者或最终用户可以提交增加新术语或修改/删除一个现有术语的意见。(2)在线编辑术语款目或处理已提交的维护建议:每种语言版本的管理员负责该种语言版本的修订工作,且每种语言版本必须与英文版本保持一致。(3)在线管理或发布新版本:由系统管理员负责,与各语种版本管理员合作完成。

该系统的输入输出设计如下。(1)允许多种数据来源:Word、SKOS、XML、RDF以及其他机器可识别格式。(2)允许多种修订方式:线上,通过在线维护系统;线下,即传统方式,修订后提交Word文件或其他机器可识别格式文件,由系统管理员转入系统。(3)实现与其他平台的数据转换和共享,提供灵活的服务方式(在版权许可下)。①输出SKOS文件:提供给ISO 5127_CLSS(发布新版本),提供下载和其他服务,面向所有用户。②输出Word文件:提供给纸本出版者。③输出XML(或其他指定格式)文件:提供给OBP或其他相关平台。④修改过的(新增/修改/删除过的)术语款目

列表可以以指定的格式从系统输出,提供给ISO进行官方审核(如有组织的投票和走工作组程序)。

总体目标是向所有用户提供搜索和浏览服务,准许授权用户根据自己选择的语言参与编辑、讨论、提交意见、修订管理及其他相关服务,并为后续进一步提供更多的网络服务做准备。

相较于从头开发一个完整的在线维护系统,采用MediaWiki实现ISO 5127术语国际标准的在线修订和语义维护可以做到开发量少和维护容易。该方案同时具备展示直观、获取方便、共享范围更广、多语言功能、可交互式协作编辑与管理、自由参与讨论、意见提交与处理、分工明确等优势,可大大提高修订工作的效率。

3 基于MediaWiki的ISO 5127在线维护系统的设计与实现

3.1 系统需求功能设计

笔者分析了ISO 5127语义在线维护的需求,结合MediaWiki,通过不断探索确定了系统的整体需求,其主要功能如下。

(1)用户管理与权限分配:这是应用系统中不可缺少的一部分。不同用户对系统功能的需求不同。出于安全性和专业性考虑,不同于一般Wiki网站的全开放,在本系统中,重要的系统功能需限定部分用户使用,并可根据需要为用户定制不同的权限。默认情况下,笔者将系统用户分为5个用户组,其中“建议用户”和“修订用户”是根据需要新设定的两个用户组,通过安装Lockdown和SimpleSecurity扩展可设置用户权限,用户和权限的具体对应信息如表1所示。

表1 基于MediaWiki的ISO 5127在线维护系统用户与权限对应表

用 户	搜索/获取/浏览历史	参与讨论	提交修订意见	新增、修改与删除	提取和处理修订意见	系统管理、开发与维护
访客	√					
登录用户	√	√				
建议用户	√	√	√			
修订用户	√	√	√	√	√	
管理员	√	√				√

(2)术语表展示:其一按照术语等级或层次结构展示,其二按照字母顺序展示。点击任一术语即进入该术语的详细信息页面,每个术语的详细页面均有针对

该术语的讨论和历史页面。

(3)术语表搜索、获取、参与讨论与浏览变更历史(可追溯性)。

(4) 术语表修订: 新增、修改和删除术语款目, 提取并处理修订意见。

(5) 修订意见提交: 参考OTCSS系统的修订规则^[5], 系统将需要采集的用户知识(修改意见)分为5种类型。①新增补术语款目(概念); ②为原术语款目增加同义词(入口词); ③修改原术语款目信息; ④原术语款目整条删除; ⑤新增关系种类。

(6) 多语言版: 可支持中文、英文、法文、德文和希腊文等多种语言的自动切换显示和编辑修订。

(7) 提供创建和发布机器可访问、可读和可转换标准的工具和过程。

系统功能可通过网站左侧导航菜单进行试用, 图1为术语表展示界面, 图2为建议用户提交意见和修订用户处理意见入口界面, 图3为修订用户修订术语操作界面。

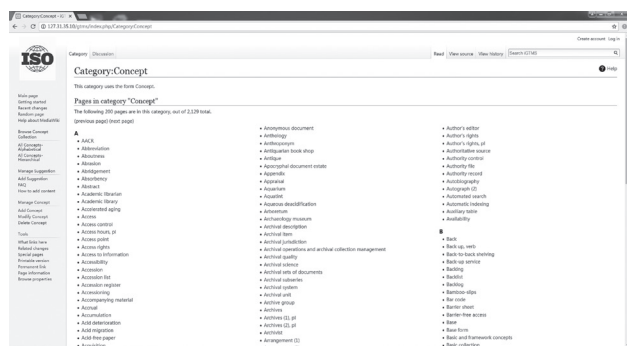


图1 术语表按字母顺序展示界面

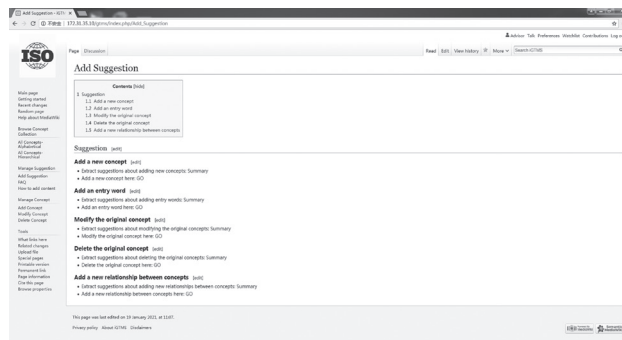


图2 建议用户提交意见和修订用户处理意见入口界面

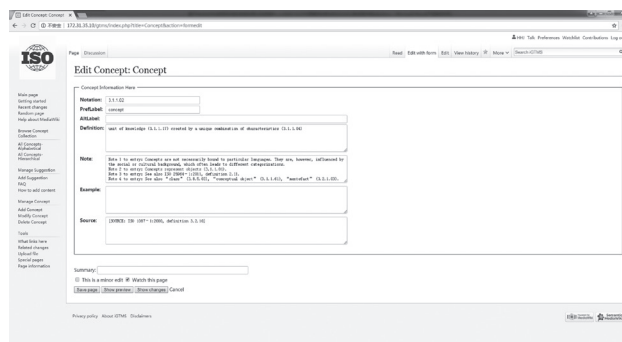


图3 修订用户修订术语操作界面

3.2 系统详细设计与实现

系统详细设计与实现的过程、方法和工具见图4所示的系统流程图。本文选取系统中较为核心的3个功能需求作为示例进行进一步的介绍。

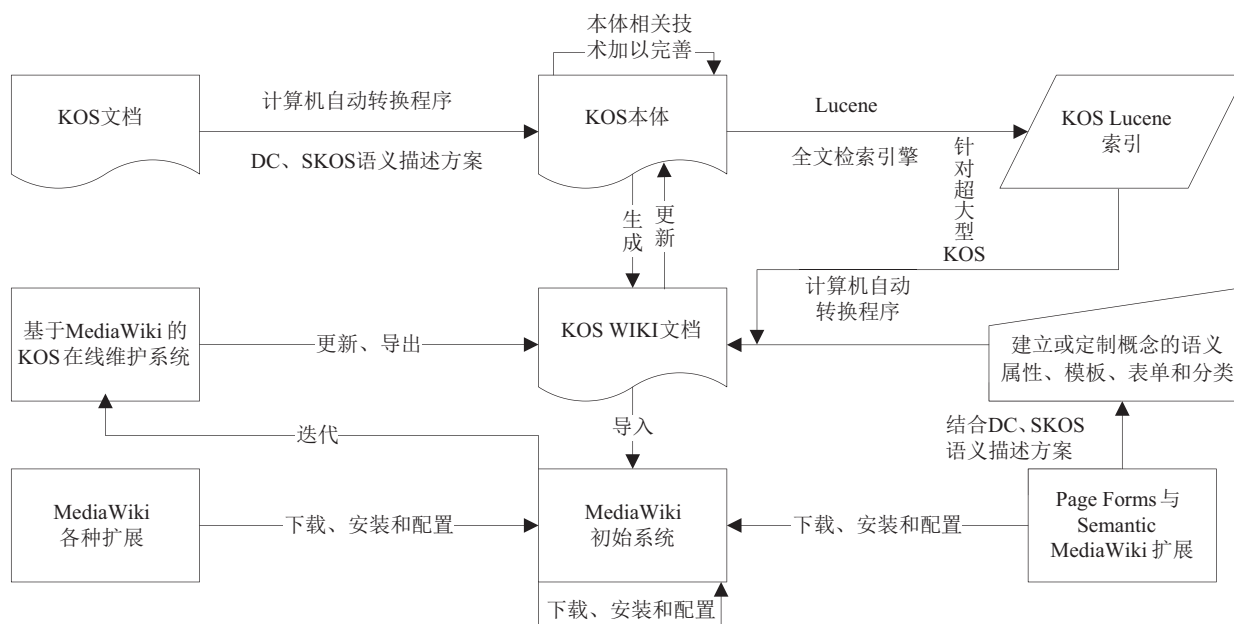


图4 系统详细设计和实现流程图

(1) 术语语义描述与MediaWiki模板定制。通过分析ISO 5127的DC和SKOS语义描述^[3], 结合MediaWiki的模版定义, 设计了一套适用于术语的表达方式。根据ISO 5127术语的skos: prefLabel值的唯一性, 使用该值作为经MediaWiki定制的术语个性化语义页面的URI或标题。通过MediaWiki的特殊页Special: SpecialPages, 首先创建ISO 5127术语款目的语义属性(关系)和数据类型, 选择是否采用列举可取值选项; 其次创建术语表管理和建议提交的页面模板(template)、表单(form)、分类(category)和入口(entrance); 最后以此类推分别创建其他语言的相应模板、表单、分类和入口。MediaWiki中ISO 5127的术语(concept)的表达方式可从该术语的“View source”中查看。

(2) 个性化需求与扩展实现。系统的个性化需求的实现基于MediaWiki丰富的扩展, 从MediaWiki下载所需的Extension, 安装与配置完成后即可实现特定的扩展功能, 甚至可以对Extension进行二次开发以实现更符合最终用户的体验需求。如术语语义化显示与编辑(下载、安装和配置PageForms与Semantic MediaWiki扩展)、多语言版(下载、安装和配置Polyglot或LanguageSelector与UniversalLanguageSelector扩展)和用户权限管理(下载、安装和配置Lockdown与SimpleSecurity扩展)等需求功能均可依此实现。

(3) 数据转换与导入。快速实现和部署基于MediaWiki的ISO 5127在线维护系统需要进行数据的转换和导入, 避免人工编辑录入的烦琐。通过厘清MediaWiki页面的特征和格式, 使用Java语言编程实现SKOS描述的ISO 5127术语向MediaWiki页面的自动转变, 然后利用MediaWiki的批量导入功能可将ISO 5127的2 126条术语全部迁移到系统。转换与导入之后, 术语的显示效果如图1所示。

限于篇幅, 其他功能请登录系统网站体验和试用。操作指引请见网站左侧菜单“How to add content”, 点击“Help about MediaWiki”可全面了解MediaWiki的各项功能。整个系统的源代码、安装步骤和详细配置都是公开的, 可以从系统网站上获取, 以促进MediaWiki在知识组织系统在线维护方面的应用和完善, 更好地实现同行之间的交流合作以达到共同进步。如需了解更多的实现细节, 请与作者联系。如有必要, 也会基于其他的KOS案例开发无版权争议的示范系统。

4 结语

本文基于MediaWiki的自由、开放、共享、可扩展和语义化等特征, 积极推进和实现了ISO 5127术语国际标准的语义在线维护系统。此系统有助于在大数据和人工智能环境下发挥人与计算机的各自优势。

本研究为来自不同ISO技术委员会(TC)、分技术委员会(SC)和所有感兴趣的团体共同维护ISO 5127术语国际标准提供了技术基础, 在2021年5月召开的第48届ISO/TC46年会上, 促进了ISO 5127 TCG(Terminology Coordination Group, 术语协调工作组, 一个常设机构, 取代之前的AHG)的成立(经由ISO/TC46 Resolution 2021/06)。我们将在下一步继续完善该系统, 为ISO 5127的维护建立完善的协调共建流程和机制。

同时, 该方案可为国内外同类型和其他类型的知识组织系统快速实现语义在线维护提供参考和借鉴。例如, 各种类型的中文知识组织系统可依据“中文知识组织系统形式化语义描述标准体系”^[5]实现语义描述, 然后可借鉴本文方案快速实现语义在线维护。

参考文献

- [1] ISO/TC 46. ISO 5127 Information and documentation- Foundation and vocabulary [EB/OL]. [2021-09-08]. <https://www.iso.org/standard/59743.html>.
- [2] 刘春燕. ISO 5127国际术语标准结构化语义分析及语义网维护研究[J]. 情报工程, 2016(6): 69-77.
- [3] 曾新红, 黄华军, 刘春燕, 等. ISO 5127的SKOS语义描述方案及其共享服务系统研究[J]. 图书情报工作, 2017, 61(21): 123-129.
- [4] 曾新红, 明仲, 蒋颖, 等. 中文叙词表本体共建共享系统研究[J]. 情报学报, 2008, 27(3): 386-394.
- [5] 曾新红, 林伟明, 黄华军, 等. 中文知识组织系统——语义描述、共建及共享服务[M]. 北京: 化学工业出版社, 2016.
- [6] 中国科学技术信息研究所. 汉语主题词表(工程技术卷)[M]. 北京: 科学技术文献出版社, 2014.
- [7] ISO/SAG on Machine Readable Standards. Final Report to the TMB [N]. 2019-12-30.
- [8] MediaWiki. MediaWiki is a collaboration and documentation platform brought to you by a vibrant community [EB/OL]. [2021-08-26]. <https://www.mediawiki.org/wiki/MediaWiki>.
- [9] Semantic MediaWiki (SMW) [EB/OL] [2021-08-27]. <https://>

- www.semantic-mediawiki.org/wiki/Semantic_MediaWiki.
- [10] Research Data Alliance (RDA). TeD-T Term Definition Tool [EB/OL]. [2021-08-26]. https://smw-rda.esc.rzg.mpg.de/index.php?title=Main_Page.
- [11] Discourse DB. The user-powered database of political commentary [EB/OL]. [2021-08-26]. https://discoursedb.org/wiki/Main_Page.
- [12] 食品百科全书 [EB/OL]. [2021-08-26]. <http://www.foodbk.com/wiki/%E9%A6%96%E9%A1%B5>.
- [13] MBA智库·百科 [EB/OL]. [2021-08-27]. <https://wiki.mbalib.com/wiki/%E9%A6%96%E9%A1%B5>.
- [14] MediaWiki. Help about MediaWiki [EB/OL]. [2021-08-26]. <https://www.mediawiki.org/wiki/Help:Contents>.
- [15] ISO. Welcome to the Online Browsing Platform (OBP) [EB/OL]. [2021-09-06]. <https://www.iso.org/obp/ui>.
- [16] ISO/TC 46/WG 4. FDIS ISO 5127 (E) revision version Annex B [N]. 2019-05-08.

作者简介

黄华军, 男, 1984年生, 硕士, 馆员, 研究方向: 知识组织、软件工程、计算机应用技术。

曾新红, 女, 1968年生, 硕士, 研究馆员, 通信作者, 研究方向: 知识组织、数字图书馆相关标准和技术, E-mail: zengxh@szu.edu.cn。

A Study on the Semantic Online Maintenance of Knowledge Organization System Based on MediaWiki: Taking ISO 5127 as an Example

HUANG HuaJun ZENG XinHong
(Shenzhen University Library, Shenzhen 518060, China)

Abstract: This paper provides a practical alternative scheme and system demonstration for knowledge organization system to realize semantic online maintenance. The features, advantages and applications of MediaWiki, which is an open source tool for Wikipedia, as well as the overall information and structural features of SKOS semantic description for ISO 5127 are studied, then an online maintenance scheme based on MediaWiki is proposed. According to the online maintenance requirements of ISO 5127, the powerful extension function of MediaWiki is adopted to realize the ISO 5127 online maintenance system, including term display, search and access, suggestion submitting, suggestion processing, real-time addition, deletion and modification, batch import and export, system logs, multilingual interface, authority management and other functions. The semantic online maintenance system implemented in this paper can basically meet the requirements of end users, and the similar systems for other types of knowledge organization systems (e.g. taxonomies, category tables, standard code tables, etc.) can be quickly implemented by drawing on the experience of this solution.

Keywords: MediaWiki; Knowledge Organization System; ISO 5127; Semantic Online Maintenance

(收稿日期: 2021-10-11)