

基于Wiki技术的标准术语库的设计与实现*

□ 王莉 梁冰 郝春云 朱晓华 白海燕 / 中国科学技术信息研究所 北京 100038

摘要: Wiki在知识协作方面的便利性,使其逐渐成为人类历史一种崭新的、高效率的知识积累方式。文章提出基于Wiki技术构建标准术语数据库建设方案,旨在集合大众力量进行知识共建;同时,利用Wiki自定义信息模板功能,在Wiki用户界面下模拟实现了数据的结构化存储与查询,以及自动抽取批量发布的功能,并在实际应用中取得了良好效果,对充分发挥标准术语的知识价值、促进标准的宣贯具有积极作用。建设实践中数据质量通过术语描述规范和术语提交审订流程控制,文章对其间可能涉及的知识产权风险进行了说明。

关键词: 标准术语, 标准术语数据库, Wiki, MediaWiki, 信息模板, 知识产权

DOI: 10.3772/j.issn.1673-2286.2011.03.010

1 引言

图书情报界一直比较重视标准化工作,早在1979年即成立了全国信息与文献标准化技术委员会(简称“文标会”,对口ISO TC46),在国家标准化管理委员会的领导下,负责制定和推广与文献和图书馆工作有关的国家标准。进入网络时代,标准规范更成为数字图书馆实现开放、互操作和集成的基础,标准选择与应用恰当与否直接关系到数字图书馆的可持续发展^[1]。

截至2010年12月,文标会组织起草国家标准70余项,其中现行标准62项^①。这些国家标准对我国图书情报工作与国际接轨,促进我国情报信息现代化有着重要意义;然而标准的宣贯推广一直举步维艰。究其原因,有标准制修订过程中工作组对需求、应用场景、相关标准等缺乏了解,导致标准本身不严谨的因素;也有用户在申请过程中没有透彻了解标准,造成难以实现的问题^[2]。解决问题的关键是对标准内容的准确理解。标准中的术语则是准确理解标准、正确传播标准、合理应用标准的保障。

“标准中的术语”(文中简称“标准术语”,区别于术语工作中的“标准化术语”),泛指标准文本中出现的术语词汇,既包含文本中术语和定义部分的内容,也覆盖了标准正文及附录。例如,《信息与文献书目数据元目录》系列标准中定义了书目工作中数据交换所需的数据元和数据元组,这些数据元和数据元组的定义都是经过严格审定的,对于标准的理解与应用具有重要价值。(注:有时候,特定学科领域内的术语条目会作为行业或国家标准结集出版,这些术语也是“标准中的术语”,但是出于对知识产权风险的考虑,这部分内容不纳入本文讨论范围。)

笔者在此背景下展开研究,详细阐述了利用Wiki技术构建标准术语数据库的方法,并在中国科学技术信息研究所2010年度预研基金的支持下得以实现,陆续收录信息与文献领域内的国家标准,取得了良好效果。

2 Wiki技术^[3-7]

Wiki是一种因互联网而产生的理念和技术,是以

*本文系中央级公益科研院所基本科研业务费专项资金中国科学技术信息研究所2010年度预研基金项目“信息与文献标准术语协作计划”(项目编号:YY-2010014)研究成果之一。
①数据检索自“国家标准化管理委员会国家标准查询系统”(http://gb.sac.gov.cn/stdinfo/servlet/com.sac.sacQuery.GjbcxServlet),检索时间:2011-01-04。

互联网开放、合作、平等、共享的精神，汇集大众信息和智慧，协同写作的超文本系统。与其他超文本系统相比，Wiki具有维护快捷、格式简单、链接方便、命名平易、可延展、开放、自组织等特点。“Wiki的魅力在于人人都能参与网页的修改，让网站内容从无到有、从粗略到完善、从孤立到互为参考不断发展下去，最终成为高质量的知识体系”^[4]。迄今为止，最为成功的Wiki项目是维基百科全书，在短短数年间规模已经超越《不列颠百科全书》，成为大量读者查找资料的有效来源。

由于Wiki在知识协作方面的便利性，其应用逐渐扩大，国外使用Wiki建立企业知识库已经很普遍，Yahoo!、NOKIA、Motorola等公司都有利用Wiki进行项目协作、知识传播的成功例子，亚马逊在线商城也在开始利用Wiki共享购物者的经验。使用Wiki技术搭建标准术语数据库基本框架，旨在集合大众力量进行知识共建。标准术语库（文中简称STWiki）以Wiki方式不断扩充，词条之间相互关联，揭示的内容比普通术语库更丰富；同时，Wiki提供了自由的交流平台，对术语条目展开的积极争辩将深化对标准的理解，更好地促进标准的制修订工作。

3 建设方案

3.1 数据来源

术语的选取与认定在术语库建设工作中占有很大的比重，一般由领域专家或词汇学家承担，自动识别与抽取技术尚未进入大规模应用。与普通的术语库相比，标准术语数据库的词条选取相对简单，其数据主要来源于两个渠道：其一是直接从标准文本中抽取，其二是借助大众参与和贡献。直接抽取的词条作为术语库的基础词条存在，主要源自标准中的术语和定义章节。此外，在标准正文和附录中也可能包含一些以定义方式显式出现的词汇，也应该直接收录。对于那些隐藏在字里行间，对准确理解标准有重要意义的词汇，则需要借助大众智慧逐步挖掘并呈现出来。对现行标准来说，这些随着更多人的参与和贡献，不断丰富与完善的内容将形成一部标准使用宝典。对制修订进程中的标准项目来说，知识共建的过程正是一个“多方参与、讨论、相互理解、相互让步的过程”^[2]，确保最终形成的标准具有普遍适用性。

3.2 数据质量控制

Wiki技术提供了高效率的知识累积方式，为了防止知识散乱局面，更有效地发挥术语价值，促进知识的传播与交流，标准术语数据库在建设过程中通过术语描述规范和术语提交审订流程对数据质量加以控制。

3.2.1 标准术语描述规范

笔者参照我国在术语工作、术语库建设、标准编制等方面的相关规定^[8-14]，对国内外有代表性的数个术语数据库进行分析，结合调研，提出一套完整的标准术语描述规范（简称STWiki Schemas）。规范中定义了3个元素集，如图1所示。

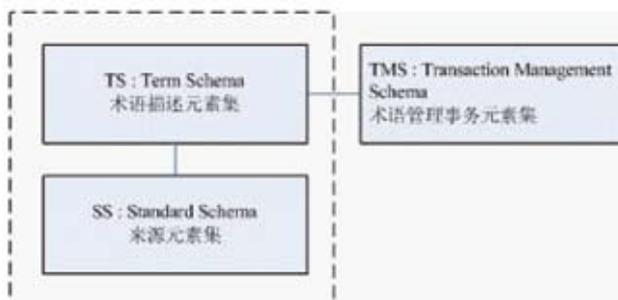


图1 标准术语著录规范中各元素集关系

(1) 术语描述元素集

术语描述元素集对术语和与术语相关的概念体系等信息进行描述，是整个规范的核心。主要数据项包括：条目标识符、主条目术语、其他语种对应词、简称、全称、同义词、拒用和被取代术语、许用术语、符号、定义、语境、概念的其他表述形式、注、专业领域、参照、语种、分类（分类表）、叙词（叙词表）、上位词、广义词、下位词、狭义词、同位词等。

(2) 来源元素集

来源元素集表示术语的出处，在标准术语数据库建设中特指标准文件，主要数据项有：来源文献类型、标准起草单位、标准名称、发布日期、实施日期、发布单位、标准号、标准分类、采用标准、采用方式、标准状态等。

(3) 术语管理事务元素集

在术语管理中一般包括创建、更新、校验、批准、注销等事务，分别对应于标准制修订流程中的征求意见、送审、报批、批准、废止等各个阶段。标准术语描述规范中采用事务类型、责任者、时间和注释四个元素组成一个描述管理事务的复合元素，适用于

标准项目执行过程中多方参与和讨论。

表1从术语描述数据项、概念及概念体系描述数据项、管理数据项和其他数据项四个方面，给出了标准术语数据库和国内外具有代表性的6个术语数据库的条目结构对照。

表1 STWiki和LEXIS、TEAM等术语库条目结构的对照^[15]

术语数据库	面向翻译			面向概念			
	LEXIS	TEAM	EURODICAUTOM	NORM-ATERM	GLOT-C	中国百科术语数据库	STWiki
描述术语的数据							
主条目术语	有	有	有	有	有	有	有
同义词	有	有	有	有	有	无	有
反义词	有	无	用注解说明	有	无	无	无
缩写	有	有	有	有	无	无	有
语法信息	有	有	用注解说明	有	无	无	无
使用地区	有	有	用注解说明	有	有	无	无
专业领域	有	有	有	有	有	有	有
注解	有	无	有	有	无	无	有
描述概念及概念体系的数据							
定义	有	有	有	有	无	有	有
语境	有	有	有	无	无	无	有
层级关系	有	有	用注解说明	有	无	无	有
主题	有	有	有	有	用概念类别说明	无	有
用于管理的数据							
术语作者	有	有	有	无	无	有	有
日期	无	有	有	有	无	有	有
出处/来源	有	有	有	有	无	有	有
其他数据项							
					用户的使用态度; 中文术语的歧义类型; 中文术语的结构格式;		
					汉语拼音		

表1中, LEXIS、TEAM和EURODICAUTOM3个术语数据库可以称为当今世界上内容最丰富、项目最完备的系统, 主要是为了翻译人员的需要而研制的。NORMATERM是法国标准化协会(AFNOR)研制的标准术语数据库, 旨在控制和存取AFNOR日益增加的术语。GLOT-C是中国学者与德国夫琅和费研究院合作研制的中文术语数据库(1986-1988), 收录了国际标准化组织从1974年到1985年期间公布的ISO-2382标准中的全部数据处理术语。从术语来源看, GLOT-C是和STWiki最接近的一个项目, 其每一个术语条目包括: 术语的索引号、中文术语、等价的英文术语、中文术语的概念类别、中文同义术语、中文多源术语、用户对术语的使用态度、术语的使用地区限制、术语的使用专业领域限制、中文术语的结构格式、中文术语的歧义类型等。中国百科术语数据库是新闻出版总署的重点科研项目, 以《中国大百科全书》为资源开发研制, 条目包括中文名称、汉语拼音、英文对应词、定义、学科领域、学科代码、来源、编辑、修订人、日期等10项^[16,17]。

3.2.2 术语提交审定流程

标准术语数据库中数据的提交审定流程借鉴吸收了维基百科的方针^[18], 主要依靠大众的监督, 通过Wiki的用户权限控制实现。首先, 系统设置管理员、行政员以及仲裁委员会, 其任命方式与职责权限参照维基百科的方针。其次, 明确系统建设目标和收录范围, 这是词条是否得到承认的基准。虽然词条的提交是自由的, 但是任何人都有权利提出异议, 而删除的权利则属于仲裁委员会。第三, 明确词条编辑规范, 术语正式定义及描述要求有出处, 严格摘自标准文本, 属于用户个人理解则在讨论页中进行。

与普通术语库相比, 标准术语数据库的正规词条直接摘自国家正式颁布的标准文本, 减少了选取上的争议。提交审定流程简单化的同时也带来了知识产权的问题。对于讨论页中的内容, 沿用维基百科“不要侵犯著作权”的方针。对于摘引国家标准的部分, 涉及的知识产权问题澄清如下^[19-22]:

(1) 标准术语数据库的知识产权问题主要指摘引标准文本这一行为涉及的著作权问题, 不涉及标准制修订过程中与专利权、商标权之间的关系。

(2) 国家强制性标准是具有法规性质的技术性规

范, 不存在著作权保护问题; 但是, 推荐性标准属于著作权法保护的范畴。

(3) 标准术语数据库以非营利的教育科研为目的, 旨在标准的推广与应用, 仅供教学或科研人员使用。

(4) 所有摘引指明出处。

(5) 为了避免对标准出版物在复制出售等方面的收益产生影响, 标准术语数据库不收录标准全文, 只是摘录术语部分内容。有些专业领域的术语直接以标准形式结集出版, 这类标准不予收录。

(6) 标准一般不会申明不许转载, 如果在收录之后, 其著作权人提出异议, 仲裁委员会将删除所有涉嫌侵权的条目。

3.3 数据库体系结构

Wiki采用页面的方式管理数据, 每个页面都有一个用标题和基本元数据标识的条目。页面本身的文字内容被存放在text表中。这种非结构化存储方式适应于人类的阅读和思维, 但是不利于机器理解, 不利于系统间的数据交换。本项目在建设过程中采用规划命名空间、规范页面分类和命名等方式梳理术语词条之间的关系, 利用自定义信息模板组织术语描述信息, 为数据挖掘打下良好基础。

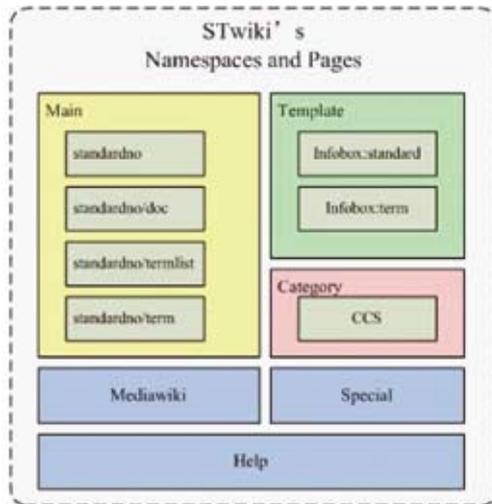


图2 命名空间及页面结构

3.3.1 命名空间

Wiki的页面标题由命名空间和页面正式标题两部

分组成，二者之间用半角冒号分隔。例如：“Help:content”中，“Help”是命名空间，“content”是页面标题。命名空间主要用于划分不同性质的页面，便于检索和管理。例如，搜索时可以通过指定名字空间缩小搜索范围；还可以为不同命名空间的页面设置独特的显示样式。在编撰一个新页面时，页面命名中命名空间前缀是可选的，如果没有指明任何前缀，该页面属于“主命名空间”（Main namespace）^[23,24]。

本项目中使用的命名空间主要包括：Main（条目页面）、MediaWiki（系统管理相关）、Special（特殊）、Template（模板）、Category（分类）和Help（帮助）。所有条目内容直接采用无前缀方式放入Main中，利用Wiki分类功能，通过添加中国标准文献分类标签对页面进行组织。

3.3.2 页面分类与命名

标准术语数据库包含六类页面：

(1) 标准详细描述页面：提供标准的详细描述信息，主要内容包括前言、适用范围、规范性引用文件、定义、包含的标准术语列表等。页面采用“正式标准号/doc”规范命名。

(2) 标准概述页面：即标准基本信息概览页面，采用Infobox standard模板控制版式与内容，描述内容对应于著录规范中的来源元素集。页面命名直接采用“正式标准号”形式。

(3) 标准术语列表页面：某项标准中包含的所有术语一览表，供浏览使用。术语可以按条目的字母顺序排列，如果标准本身对其概念层次作了规定，也可以按规定层次排列。页面命名采用“正式标准号/termlist”形式。

(4) 标准术语描述页面：术语详细描述信息，采用Infobox term模板控制版式与内容，描述内容对应于著录规范中的术语描述元素集。页面采用“正式标准号/术语名”规范命名。

(5) 工具页面：支持词条撰写、数据抽取等系统功能，主要包括自定义的Infobox模板、结构化数据抽取与发布、RDF查询等。页面采用Wiki特殊页面命名方式，例如，模板的定义命名为“模板:infobox 模板实名”。

(6) 帮助页面：提供系统使用方面的帮助信息，采用“帮助:帮助主题”方式命名，统一纳入“帮助:目录”下管理。

3.3.3 自定义信息模板

标准术语数据库中提供了standard和term两个自定义信息模板，分别对应著录规则中的来源元素集和术语描述元素集。

(1) Template: infobox standard定义

```

{{Infobox standard
|standard_no      =
|title            =
|status           =
|year_started     =
|year_ended      =
|editors          =
|base_standards  =
|relationship     =
|related_standards =
|abbreviation    =
|domain          =
|website         =
}}
```

standard模板采用12个变量对标准进行描述，其中relationship和related_standards成对出现，用来表示当前标准和其他标准之间的关系（如：等同采用、修改采用、非等效标准）。图3是模板使用效果，关于模板的详细用法参见页面“Infobox standard/doc”。



图3 GB/T 19688.5-2009概述页面^②（系统截图）

^②国家标准《信息与文献 书目数据元目录 第5部分：编目和元数据交换用数据元》（GB/T 19688.5:2009）概述页面。

```

(2) Template : infobox term定义
{{Infobox term
|main_entry_term          =
|transliterated_form      =
|abbreviated_form_of_term =
|full_term                =
|synonym                 =
|deprecated_and_superseded_term =
|admitted_term           =
|symbol                  =
|definition               =
|context                 =
|nontextual_illustrations =
|note                    =
|subject_field            =
|cross_reference          =
|language_code            =
|URI                     =
|standard_website         =
}}
```

以上17个变量分别从表述形式、概念、定位与关联等三个方面对标准术语进行描述。其中cross_reference（相互参照）、URI（条目唯一标识）、standard_website（来源标准唯一标识）用来定位与关联，这些变量是构建概念空间的基础。图4是模板使用效果，关于模板的详细用法参见页面“Infobox term/doc”。

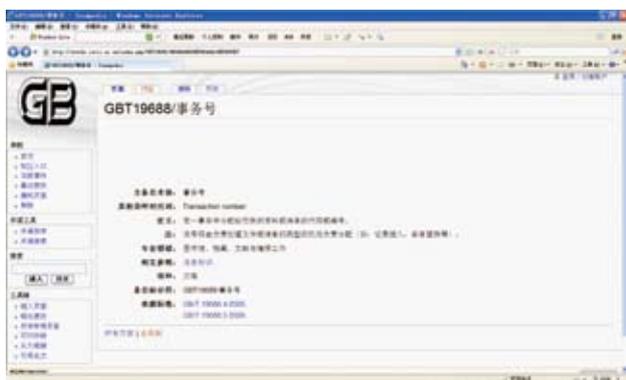


图4 GBT19688/事务号^③（系统截图）

4 系统实现

标准术语数据库以MediaWiki 1.15.1为基础，采用Apache2.2.11+MySQL5.1.36+PHP5.3.0框架，在Windows平台上构建成功。系统体系结构图如图5。

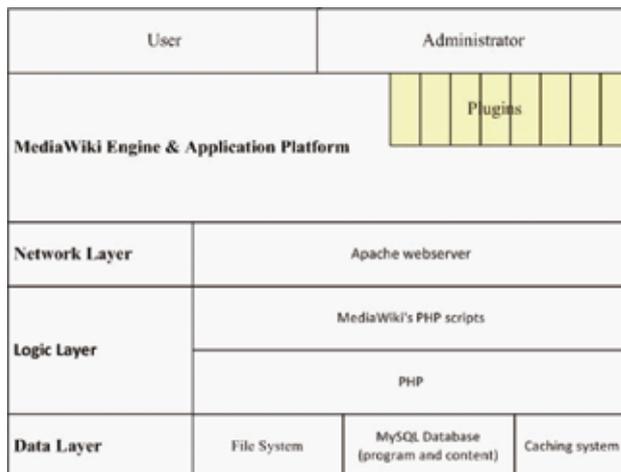


图5 体系结构图^[25]



图6 系统主要功能模块

图6展示了系统的主要功能模块，在实现全部Wiki功能的基础上，增加了术语发布和搜索。系统实现过程中的关键在于：如何在Wiki的非结构化数据中实现结构化数据的存储与查询功能^[26,27]。简单地说，解决该问题的基本方法有两个。

方法一：底层采用关系型数据库的表存储结构化数据，通过编程实现Wiki页面上对数据的动态操作，由Wiki页的用户界面提供数据。即，写Wiki页面时，数据同时以结构化方式存储在数据库表里；读Wiki页面

^③国家标准《信息与文献 书目数据元目录》(GB/T 19688)中“事务号”的描述页面。

时,从数据库表里提取结构化数据动态显示。

目前,可以从MediaWiki扩展库中查到很多可供利用的扩展程序(主要在Category:Parser_function_extensions类目下),如External_Data、DataInvoker和DataTable^④等。自己编写扩展程序也十分简单,利用MediaWiki的各种内部hook、类和方法将特定功能无缝嵌入MediaWiki程序。

方法二:不改变Wiki的数据管理方式,而是通过解析页面信息的手段动态获取结构化数据。

Web页面信息的自动提取是数据挖掘领域里一个重要的研究方向,虽然国内外已经有很多方法,但是实现起来比较复杂,而且准确率也比较低。其难点在于规则的自动发现与获取。如果已知页面的固定框架,可以绕过语料训练、机器学习的步骤直接采用模板方法准确抽取数据。

标准术语数据库在建设初期即对命名空间、页面名称、描述内容等作了详细规划,采用自定义信息模板实现对显示格式的控制,更重要的是,这种固定的描述框架为信息的提取打下良好基础。系统实现中采用方法二,巧妙利用Wiki自身的模板功能,通过自定义方式存储词条的结构化信息。术语发布实际是解析信息模板,批量抽取结构化数据的过程;术语查

询则利用DynamicPageList2^⑤扩展程序实现。

5 结语

标准术语数据库第一阶段以文标会第四分会组织实施的国家标准制修订项目为术语来源,基于Wiki技术建设自由开放的标准术语协作平台,集合大众力量进行知识共建,词条之间相互关联,揭示的内容比普通术语库更丰富;同时,项目巧妙利用Wiki自定义信息模板功能,在Wiki用户界面下模拟实现了数据的结构化存储与查询,以及自动抽取批量发布的功能,取得良好的效果,对充分发挥标准术语的知识价值、促进标准的宣贯具有积极作用。

需要提到的是,信息模板在Wiki中真正的作用只是显示效果的控制,对数据质量并没有强制约束力。因此,利用信息模板实现结构化数据存储、查询与发布的方法,虽然实现起来简单快捷,但并不完善。严谨的质量控制需要从词条编撰规范、提交审核流程和技术手段三个方面同时进行。相关的词条编撰规范和提交审核流程已经形成,更严谨的技术控制手段则是本项目的下一步工作。

参考文献

- [1] 张晓林,鞠文红,李珍. 数字图书馆标准规范发展趋势[R/OL]. <http://202.113.20.229/TPI/SYSASP/include/docs/%E6%95%B0%E5%AD%97%E5%9B%BE%E4%B9%A6%E9%A6%86%E6%A0%87%E5%87%86%E8%A7%84%E8%8C%83%E5%8F%91%E5%B1%95%E8%B6%8B%E5%8A%BF.pdf>.
- [2] 孙卫,赵悦. 图书馆信息化标准工作评析[J]. 数字图书馆,2009(7/8):33-37.
- [3] Wiki技术及应用[EB/OL]. [2010-12-17]. <http://www.hudong.com/Wiki/Wiki%E6%8A%80%E6%9C%AF%E5%8F%8A%E5%BA%94%E7%94%A8>.
- [4] 天下维客[EB/OL]. [2010-12-17]. <http://www.allWiki.com/Wiki/Wiki>.
- [5] 徐宏宇. Wiki-开放式的资源共建模式[EB/OL]. [2010-12-17]. <http://www.istis.sh.cn/list/list.aspx?id=4306>.
- [6] 陆泉,邓晶,胡慧丽. 基于Wiki的数字图书馆信息服务平台研究[J]. 图书馆论坛,2010(2):13-16.
- [7] 郭桂英,毛春华,杨美清. WIKI应用案例评述——来自Harvey Mudd College和东京慈惠会医科大学的研究[J]. 中国教育信息化(高教职教),2008(4):80-82.
- [8] GB/T.6-88, 标准化工作导则术语标准编写规定[S].
- [9] GB/T1011-88, 确立术语的一般原则与方法[S].
- [10] GB/T10112-959, 术语工作原则与方法[S].
- [11] GB/T13725-2001, 建立术语数据库的一般原则和方法[S].
- [12] GB/T15237.1-94, 术语数据库开发用文件编制指南[S].
- [13] GB/T15387.2-94, 术语数据库开发指南[S].
- [14] GB/T15625-1995, 术语数据库技术评价指南[S].
- [15] 现代术语学[EB/OL]. [2010-01-20]. <http://www.asic.org.tw/Content/Publication5/E07-0000.htm>.
- [16] 冯志伟. 计算机辅助术语研究浅谈[J]. 术语标准化与信息技术,2002(3):11-15,24.
- [17] 王渝丽. 术语学与百科全书的编撰[J]. 科技术语研究,2005,7(1):27-28.
- [18] 维基百科:方针与指引[EB/OL]. [2011-01-04]. <http://zh.wikipedia.org/zh-cn/Wikipedia:RULES>.
- [19] 吴汉东. 论著作权作品的“适当引用”[EB/OL]. [2011-01-04]. <http://www.fengxiaqingip.com/ipteseluntan/luntan3/zhuzuoxx/20090221/3602.html>.
- [20] 刑造宇. 标准的知识产权管理策略刍议[J]. 浙江工商大学学报,2001(5):26-32.
- [21] 国际标准化组织(ISO)的知识产权保护[EB/OL]. [2011-01-04]. <http://www.tc357.com.cn/news.asp?id=150>.

^④http://www.mediaWiki.org/Wiki/Extension:External_Data.
^⑤<http://www.mediaWiki.org/Wiki/Extension:DataInvoker>.
<http://www.mediaWiki.org/Wiki/Extension:DataTable>.

- [22] 标准中的知识产权[EB/OL]. [2011-01-04]. <http://hi.baidu.com/myllaw/blog/item/8d2075ee35de0afab3fb9508.html>.
- [23] MediaWiki的名字空间[EB/OL]. [2010-12-17]. <http://www.allwiki.com/wiki/%E5%90%8D%E5%AD%97%E7%A9%BA%E9%97%B4>.
- [24] Wikipedia:Namespace[EB/OL]. [2010-12-20]. <http://en.wikipedia.org/wiki/Help:Namespace>.
- [25] Manual:MediaWiki architecture [EB/OL]. [2010-12-20]. http://www.mediawiki.org/wiki/Manual:MediaWiki_architecture.
- [26] 在MediaWiki的非结构化数据中部分实现结构化数据库功能[EB/OL]. [2010-12-20]. <http://www.jamesqi.com/%E5%9C%A8MediaWiki%E7%9A%84%E9%9D%9E%E7%BB%93%E6%9E%84%E5%8C%96%E6%95%B0%E6%8D%AE%E4%B8%AD%E9%83%A8%E5%88%86%E5%AE%9E%E7%8E%B0%E7%BB%93%E6%9E%84%E5%8C%96%E6%95%B0%E6%8D%AE%E5%BA%93%E5%8A%9F%E8%83%BD>.
- [27] 在MediaWiki系统中查询结构化数据库[EB/OL]. [2010-12-20]. <http://www.jamesqi.com/%E5%9C%A8MediaWiki%E7%B3%BB%E7%BB%9F%E4%B8%A D%E6%9F%A5%E8%AF%A2%E7%BB%93%E6%9E%84%E5%8C%96%E6%95%B0%E6%8D%AE%E5%BA%93>.

作者简介

王莉，硕士，中国科学技术信息研究所高级工程师，全国信息和文献标准化技术委员会技术协作分技术委员会委员。研究课题：数字图书馆。E-mail: wangli@istic.ac.cn

梁冰，博士，中国科学技术信息研究所高级工程师。E-mail: liangb@istic.ac.cn

郝春云，硕士，中国科学技术信息研究所高级工程师。E-mail: chyhao@istic.ac.cn

朱晓华，学士，中国科学技术信息研究所工程师。E-mail: zhuxh@istic.ac.cn

白海燕，硕士，中国科学技术信息研究所副研究员。E-mail: bhy@istic.ac.cn

Design and Implementation of a Wiki-based Standard Terminological Database

Wang Li, Liang Bing, Hao Chunyun, Zhu Xiaohua, Bai Haiyan / Institute of Scientific & Technical Information of China, Beijing, 100038

Abstract: Wiki provides an extremely fast and efficient way to collaborate and communicate knowledge. This paper puts forward a design model of a standard terminology database based on wiki technology, for the purpose of collaborative knowledge. Using the Wiki's feature of infobox templates, the design model achieves structured storage and query functions in the Wiki user interface, and supports batch extracting and publishing data. A standard terminology database is constructed using the design model and achieves good results. The terminology database improves the knowledge value of standard term, and has a positive effect on the propaganda and Implementation of standards. Means of data quality control include STWiki schemas and term submission and review process. This paper describes the intellectual property risks involved in the terminology database.

Keywords: Standard terms, Standard terminological database, Wiki, Mediawiki, Infobox template, Intellectual property

(收稿日期: 2010-12-20)