

# 语义wiki在词表编制中的应用探讨\*

范炜 陈圭璋

(四川大学信息管理技术系, 成都 610064)

**摘要:** 通过分析传统词表的发展瓶颈和语义wiki的特征, 讨论了语义wiki用作词表技术平台的可行性与具体实现思路。最后指出语义wiki作为一种技术手段, 它为传统词表研究与当前的语义网应用搭建了一个互通的桥梁, 使得传统词表能够以较低的成本和较快的方式融入到当前的信息环境中来, 从而发挥其作为组织工具和术语资源的双重作用。

**关键词:** 语义wiki; 词表开发; 术语服务

**中图分类号:** G254

**DOI:** 10.3772/j.issn.1673—2286.2014.04.003

## 1 传统词表在新时期的定位与发展问题

传统意义的词表是指运用各种情报检索语言(如分类法、主题法等)编制的专业性或综合性的检索工具, 是文献资料组织和管理的依据。在数字化网络环境中, 传统词表遭遇了信息环境变革之下的发展瓶颈, 自身发展严重滞后, 无法满足数字化网络时代对主题内容和形式的组织要求。就国内词表发展现状而言, 许多词表还停留在印本阶段, 以当前的视角进行重新审视, 对于长期积累下的词表在新时期具有两方面的作用: 组织工具和术语资源。虽然信息媒介数字化了, 但传统词表蕴含的分类法、主题法等基本组织方法仍然可以用作网络信息资源组织的参考。另一方面, 各种专业性和综合性的词表收集整理了相关术语概念及语义关系, 已经形成丰富的术语资源。本体工程可以利用词表中的术语资源来搭建本体, 而不必从零开始, 从而节约开发成本和提高效率。就组织架构的复杂性而言, 本体是目前各种知识组织系统中最复杂的一种, 规范化的定义为语义推理的实现奠定了良好的基础, 但同时我们也意识到本体构造是最耗费人力物力的。如果以本体为目标, 将各种词表进行模型改造、细化语义关系, 继而一次进化为一个个体知识库是不太现实的。与本体相比, 虽然传统词表结构在主题揭示、语

义关系描述方面不甚精确, 但其包含的术语资源是值得继承与拓展的。一种策略是在不改动词表原有数据结构的基础上, 先将存在于纸本介质的各种词表利用现有的技术手段先进行编码与表征, 使其可以实现术语资源的共享与交流, 同时也解决了词表自身的维护与管理工作。在此基础上根据不同的组织功能需求, 采用本体工程的方法来对这些术语资源进行“取材”, 进而构建各种领域本体。

如何发挥词表作为组织工具和术语资源这两方面的作用, 与词表自身的发展密切相关。目前词表发展的主要障碍是缺少有效支撑词表开发与维护的技术平台。本文从传统词表自身发展的角度出发, 基于新建术语资源的技术工具和已有词表的管理维护问题的双重考虑, 结合当前的技术环境, 讨论了利用语义wiki作为词表开发与管理技术平台的实现思路。

## 2 语义wiki分析

wiki是Web2.0发展浪潮中的中坚力量。wiki本身可以看作是一种信息组织、协作与交流的模式, 突出了“人人贡献、受益人人”的哲学思想。维基百科(wikipedia)是wiki模式运作的成功案例。从使用者的角度看, 用户只须掌握少量的wiki语法即可参与内容

\* 本研究得到四川大学中央高校基础科研业务经费项目“关联数据集描述与发现服务研究”(编号: SKQ201204)资助。

创建、链接、编辑与管理等各种wiki活动,通过内容单元错综复杂的链接,形成了一个庞大的知识网络。这种简单易用的低门槛造就了今天维基百科的巨大成功,分布在世界各地的人们不断向维基百科贡献知识,应该说以用户为中心的思想在此得到了充分体现。转向信息组织和技术角度来看,wiki的组织模式是一种松散的网状结构,其底层的数据结构与编码在一定程度上讲是非结构化和语义化的,如何运用语义网技术对wiki底层架构进行改造和管理,使其加强在机器可读可理解层面实现资源的共享与重用成为当前语义网应用的一个新兴研究主题,即语义wiki的研究。关于语义wiki的研究始于2005年,最早成立了[swigig]邮件列表,截至目前已经举办五届语义wiki研讨会<sup>[1]</sup>,其主题分别为“从wiki走向语义”(2006)、“基于wiki的知识工程”(2006)、“语义的wiki之路”(2008)、“语义wiki网络”(2009),关联数据与人(2010)。

语义wiki可以看作是一种语义网的架构模式,它结合了wiki与语义网两者的特征,一方面保留和扩展wiki的简单易用性,为用户提供良好的使用体验,促进用户的积极参与;另一方面,基于wiki平台运用语义网技术对wiki底层架构进行改造和管理,使其加强在机器可读可理解层面实现资源的共享与重用。这样既满足了非技术人员(如普通大众、主题专家)参与内容创建与共享,又实现了结构化语义数据表示与存储。

目前语义wiki的一个重要应用是用来构建领域协作式本体开发平台,如美国国家癌症研究所NCI的caBIG@知识中心的语义wiki引擎LexWiki是基于Semantic MediaWiki的定制开发,除了提供一套完整的在线协作式本体开发平台,还实现了与Protégé的有效整合<sup>[2]</sup>。BiomedGT即是基于LexWiki的一个部署实例,其术语资源包含NCI叙词表和纳米技术、生物医学等专业领域术语,为生物医学专家提供了一个本体开发与维护平台,在此基础上提供了一些术语搜索、导航与浏览功能<sup>[3]</sup>。传统词表可以借鉴语义wiki用作协作式本体开发平台的应用经验来实现自身作为一种知识资源的技术发展路线。语义wiki作为词表开发与维护平台具有以下几点优势:

#### 1) 合作机制

多个领域专家可以协同进行开发,社群(community)驱动的词表编制与维护模式,有效提高了开发效率。

#### 2) 开源机制

与商业化的词表管理软件和解决方案(如MultiTes、Synaptica等)相比,开源的语义wiki引擎有效降低了词表系统开发的技术成本。

#### 3) 易用性

继承wiki方便用户贡献内容的界面交互模式,对非技术人员屏蔽了底层的技术复杂性,领域专家通过简单的表单形式提交、编辑术语单元。

#### 4) 网络发布

利用语义wiki平台开发的词表资源可以直接通过该平台实现对外发布,普通用户使用语义wiki提供的搜索与导航功能实现对词表资源的直接访问。

#### 5) 缩短修订周期

利用语义wiki的管理机制与社群力量实现词表的快速更新与修订,使得词表包含的术语资源处于不断更新变化之中,永远的beta版本。

#### 6) 词表资源共享与重用

语义wiki可以支持RDF、OWL、SKOS等多种编码格式,并可实现词表资源的导入与导出,这使得词表资源共享与集成变为可能,将词表系统转变为一个开放的术语资源系统。

### 3 基于语义wiki的词表平台实现方案

以上谈及的wiki和语义wiki是指架构的模式(pattern),而具体的语义wiki实现方案则是多种多样,现有的语义wiki引擎有Semantic MediaWiki、AceWiki、IkeWiki、Ontoprise公司的SMW+等。MediaWiki是维基百科采用的wiki引擎,应用广泛且发展较为成熟,Semantic Mediawiki(以下简称为SMW)是Mediawiki的一个扩展,已经存在一些应用SMW进行知识库的构建项目<sup>[4]</sup>,如Semanticweb.org、MyWikiBiz、Sydney Directory、MIT的SIMILE项目wiki等。鉴于MediaWiki引擎应用的广泛性和扩展的灵活性,加上存在大量的开发资源可供选用,因此本文选用Semantic Mediawiki作为词表开发平台的语义wiki引擎展开讨论。以下是基于Mediawiki平台的语义Wiki技术框架示意图,如图1所示,其思路是以Semantic Mediawiki(Mediawiki的一次扩展)作为语义wiki核心引擎,集成Mediawiki大平台上的一些现有扩展资源(如Semantic Forms等),同时在Semantic Mediawiki基础上进行二次扩展(如改进界面用户体验的Halo扩展等)。本文构建过程中涉及的扩展除了SMW核心之

外,还包括Semantic Forms、Semantic Drilldown、Halo等,这里先简单介绍一下这些扩展。Semantic Forms扩展提供简单易用的表单界面,方便用户向wiki页面增加或编辑语义数据<sup>[5]</sup>。Semantic Drilldown扩展提供分面浏览界面,对语义属性进行分类和过滤,帮助用户在wiki中查看语义数据<sup>[6]</sup>。Halo扩展优化了wiki界面的用户体验,扩展了wiki的导航、搜索与交互行为,提供自动补全等功能<sup>[7]</sup>。

### 3.1 词表编制平台基本架构

在语义wiki平台上进行词表开发,首先需要采用

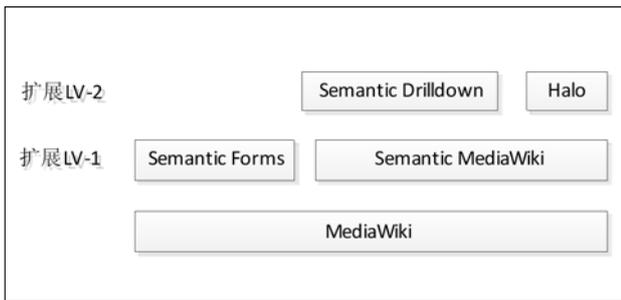


图1 SMW技术架构示意图

SMW的基本语法对词表结构进行建模,这部分可以由知识工程师与领域专家合作完成。具体工作是利用SMW的属性定义术语关系,结合语义模板和语义表单实现词表编制平台的基础架构。

SMW采用RDF的三元组模型“主语-谓语-宾语”进行属性描述与关系定义,包含三条基本语法<sup>[8]</sup>:

- 定义一个RDF三元组 `[[property-name::value]]`
- 定义属性 `[[has type::type-name]]` 属性类型包括String、Page、Boolean、Number等
- 定义分类Category `[[category:XXX]]`

这三条基本语法可以用于对各种词表结构进行建模,以下结合叙词表结构演示构建叙词表的一种基本思路,具体构建流程如图2所示。

传统的叙词表一般包括“用代属分参族”六种关系,分别以USE\UF\BT\NT\RT\TT来表示。在语义wiki中定义每一个主题词(包括叙词与非叙词)为一个Page(SMW标准属性类型)。以下是在BT的属性页面输入的代码示例,其他关系定义方式与此相同。

```
/*--
```

BT“属”即上位类关系,用于从下位类指向上位

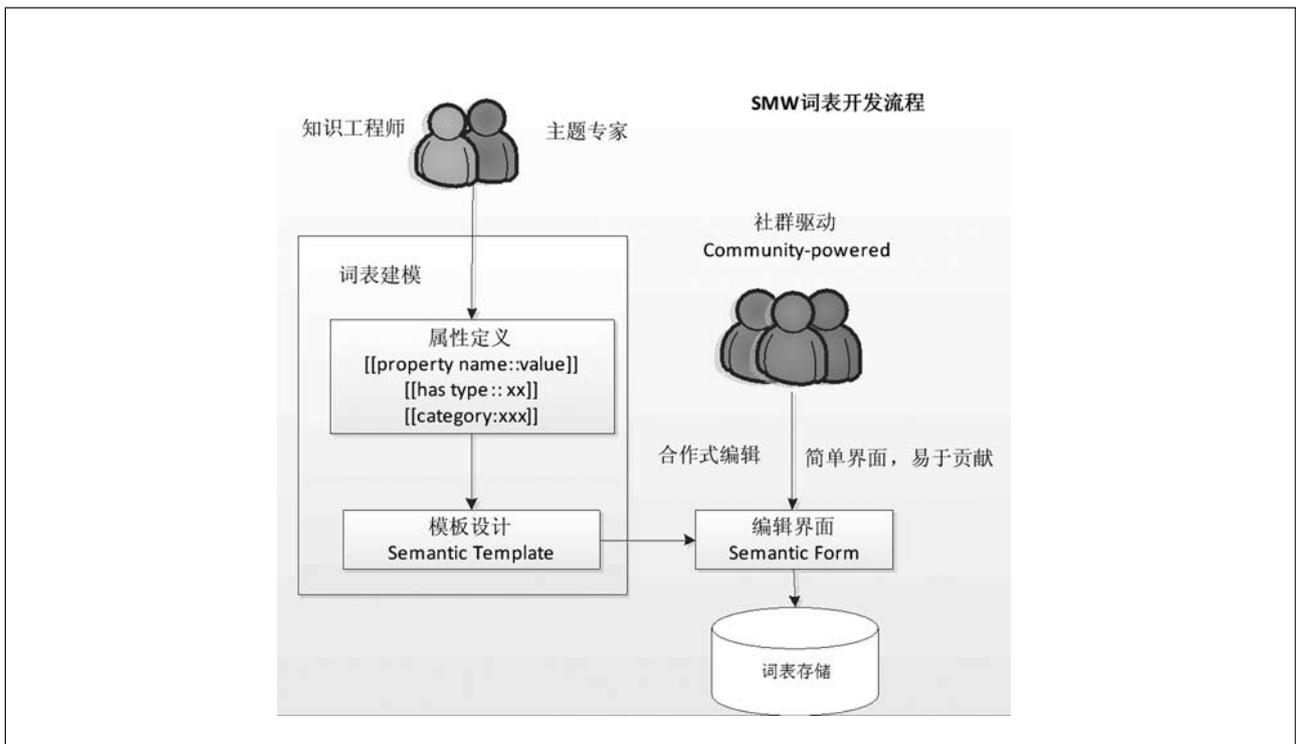


图2 基于SMW的词表开发流程图

类。BT与NT关系互为反参照。

BT关系[[has type::page]]

--/

构建好叙词表结构后，接下来需要使用模板功能来定义主题词页面，这里对前期定义的六种关系根据主题词本身的需要作以下限定：BT\NT\RT\UF关系允许多个取值，USE/TT只允许单一取值。在词表结构和主题页面模板基础上需要一个面向最终用户可直接参与编辑的界面。这里使用语义表单（Semantic Forms）设计出一个简单输入界面，方便领域专家与普通用户提交、编辑和管理词表中的主题词。最终得到的主题词页面主要分为两个部分：上半部分是按照模板定义的相关属性字段；下半部分是一个关于该主题词的事实栏（Facts box），栏内包含所有的属性和取值的动态链接，点击不同的属性与取值被链接到相关的页面。表1包括了主题词页面模板核心代码、主题词表单输入页面截图、最终主题词展示页面。

这里SMW采用RDF和OWL语法对主题词页面进行描述，上图主题词事实栏的右上角有一个RDF Feed，这是将该主题词页面的事实以RDF进行描述和编码，可供其他服务调用。SMW中允许使用模板功能来设计输出不同的编码格式，对于传统词表编码格式而言，SKOS是目前较为理想的选择。SKOS针对各类知识组织系统KOS的一个RDF应用，与OWL兼容，它为传统词表在语义网中的发展提供一种轻量级的描述机制，值得考虑采用。

### 3.2 搜索与浏览

SMW内置了简单的语义搜索功能，在Special:ASK页面的搜索框输入查询条件即可实现基于属性的语义检索。例如查询族首词“出版物”下包含哪些主题词，输入[[TT::出版物]]点击查询即可得到TT属性取值为“出版物”的所有关键词。同时SMW还支持内查询

主题词页面的模板定义	语义表单提交页面
<pre> &lt;pre&gt; {{Term  BT=  RT=  NT=  USE=  UF=  TT= }} &lt;/pre&gt; &lt;noinclude&gt;&lt;includeonly&gt; {  class="wikitable" !BT  {{#arraymap:{{{BT}}},x  [BT::x]}}  - !RT  {{#arraymap:{{{RT}}},x  [RT::x]}}  - </pre>	<div data-bbox="726 1056 1340 1498"> <p><b>Form:Term</b></p> <p>这是主题词提交表单。领域专家和一般用户可以通过这个表单页面非常容易地提交和修改主题词。请输入要添加的主题词，然后点击右侧按钮。</p> <p><input type="text"/> <input type="button" value="Add or edit"/></p> <hr/> <p>special</p> <p><b>Add Term: 出版物</b></p> <p>主题词</p> <p>BT: <input type="text"/></p> <p>RT: <input type="text"/></p> <p>NT: <input type="text"/></p> <p>USE: <input type="text"/></p> <p>UF: <input type="text"/></p> <p>TT: <input type="text"/></p> <p>Free text:</p> <p><input type="text"/></p> <p>Summary: <input type="text"/></p> <p><input type="checkbox"/> This is a minor edit <input type="checkbox"/> Watch this page</p> <p><input type="button" value="Save page"/> <input type="button" value="Show preview"/> <input type="button" value="Show changes"/> <input type="button" value="Cancel"/></p> </div> <p>主题词页面（关系中的所有主题词都可点击，链接到对应的主题词页面）</p> <div data-bbox="726 1541 1340 1983"> <p>page discussion edit history delete move protect watch refresh</p> <p><b>期刊</b></p> <p>BT 报刊</p> <p>RT 办刊, 创刊</p> <p>NT 定期刊物, 不定期刊物</p> <p>USE <span>△</span></p> <p>UF 杂志, 刊物</p> <p>TT 出版物</p> <hr/> <p><b>Facts about 期刊</b> <span>Ⓞ</span> <span>RDF Feed</span></p> <p>BT 报刊 <span>+</span> <span>🔗</span></p> <p>NT 定期刊物 <span>+</span> <span>🔗</span>, and 不定期刊物 <span>+</span> <span>🔗</span></p> <p>RT 办刊 <span>+</span> <span>🔗</span>, and 创刊 <span>+</span> <span>🔗</span></p> <p>TT 出版物 <span>+</span> <span>🔗</span></p> <p>UF 杂志 <span>+</span> <span>🔗</span>, and 刊物 <span>+</span> <span>🔗</span></p> <p>USE <span>△</span></p> <hr/> <p>Category: 中国分类主题词表</p> </div>

(Inline Queries), 在页面中直接动态集成查询结果, 基本语法是`{{#ask:arg1|arg2|...}}`。这项功能可以根据不同需求对术语资源进行组合输出与呈现。

Semantic Drilldown扩展结合了wiki的分类category功能实现语义数据的浏览(Special:BrowseData页面)和在分类页面上设置过滤器进行搜索限定。在词表系统中可以采用此扩展实现对主题词的各角度归类(分面浏览)与类目内搜索限定, 作为一种辅助查询功能。

Halo扩展集中于用户界面端的优化和改进, 在wiki页面右侧提供一个工具栏帮助用户以更直观的方式查看、注解和修改页面内容, 还提供本体浏览、改进的语义搜索界面、异步实时交互响应等。这些改进对降低语义wiki使用门槛具有重要意义。在词表平台中加入Halo扩展使得普通用户对词表的主题词编辑更加容易, 浏览和搜索主题词更加方便和直观。

以上使用到的SMW扩展主要用于增强词表平台基本服务功能和优化用户体验方面, 对于具体的功能需求还可以考虑集成其他扩展或自主进行开发。

### 3.3 词表修订与版本管理

词表的修订一直以来是一项费时费力的工程, 词表规模越大, 修订周期越长, 耗费就越大。使用语义wiki开发的词表平台承袭了wiki大众参与建设的思想和版本管理机制, 将词表的修订与维护工作变为一个可控的开放平台, 通过社群驱动的方式改进词表修订方式, 这样可以有效缩短修订周期, 使得词表迅速响应主题事物的发展变化, 及时进行调整。这一部分的具体实施以wiki原有的版本管理机制作为参照进行展开。首先在原有词表修订规则的基础上进行部署, 词表编制委员会仍然作为修订的核心, 一方面通过词表平台的协作机制如使用讨论页进行主题词修订意见协商, 另一方面对于用户贡献的主题词进行审核, 以便后来纳入术语资源库中。对于主题词术语的修订则通过历史页面进行控制。这是利用wiki的协作和编辑制度实现词表修订和版本管理的基本思想, 对于不同的词表平台还可以进行权限细分, 制定出更加细致的编辑制度。例如这里只允许普通用户贡献主题词、在讨论页发表修订意见, 而不允许在已有主题词上进行编辑操作, 对已有主

题词的修订只能由词表编制委员会来操作。这样的制度既维护了已建成词表的完整性和一致性, 又不失词表的开放性和动态性。

## 4 结语

以上对语义wiki在词表编制中作为技术平台的部署进行了初步的尝试, 为词表提出一条切实可行的发展思路。基于语义wiki的词表平台作为一种技术手段, 它为传统词表研究与当前的语义网应用搭建了一个互通的桥梁, 使得传统词表能够以较低的成本和较快的方式融入到当前的信息环境中来。开放、协作式词表开发平台作为一个术语资源开发与服务的基础架构为开展术语注册、词表仓储、语义搜索与关联数据融合等一系列术语增值服务奠定了基础。因此客观地预见, 传统词表作为术语资源的重要基石以语义wiki作为技术支撑融入到数据的网络(Web of Data)中去, 最终促进知识资源的共享与利用。

### 参考文献

- [1] SemWiki.org: The Semantic Wiki Community [EB/OL]. [2014-03-31]. <http://www.semwiki.org/>.
- [2] NCI caBIG® Knowledge Center LexWiki [EB/OL]. [2014-03-31]. <https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/LexWiki>.
- [3] BiomedGT Collaborative Ontology Development Wiki [EB/OL]. [2014-03-31]. [http://biomedgt.nci.nih.gov/wiki/index.php/Main\\_Page](http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page).
- [4] Sites using Semantic MediaWiki [EB/OL]. [2014-03-31]. [http://meta.wikimedia.org/wiki/Sites\\_using\\_Semantic\\_MediaWiki](http://meta.wikimedia.org/wiki/Sites_using_Semantic_MediaWiki).
- [5] Semantic Forms Extension [EB/OL]. [2014-03-31]. [http://www.mediawiki.org/wiki/Semantic\\_Forms](http://www.mediawiki.org/wiki/Semantic_Forms).
- [6] Semantic Drilldown Extension [EB/OL]. [2014-03-31]. [http://www.mediawiki.org/wiki/Extension:Semantic\\_Drilldown](http://www.mediawiki.org/wiki/Extension:Semantic_Drilldown).
- [7] Halo Project [EB/OL]. [2014-03-31]. <http://www.projecthalo.com>.
- [8] Semantic MediaWiki [EB/OL]. [2014-03-31]. [http://semanticmediawiki.org/wiki/Semantic\\_MediaWiki](http://semanticmediawiki.org/wiki/Semantic_MediaWiki).

## 作者简介

范炜，男，1981年生，管理学博士，四川大学公共管理学院信息管理技术系讲师，研究方向：信息组织与信息检索。E-mail: fanwscu@163.com。  
陈圭璋，男，四川大学公共管理学院信息管理技术系本科生。

### The Use of Semantic Wiki as Vocabulary Development Platform

FAN Wei CHEN GuiZhang

(Department of Information Management Technology, School of Public Administration, Sichuan University, Chengdu 610064, China)

Abstract: This paper analyzes traditional vocabulary development bottleneck and introduces the characteristics of semantic wiki, then discusses a semantic wiki way to deploy vocabulary development platform. Finally, it points out that semantic wiki as a technology tool builds a bridge between traditional vocabulary research and current semantic web applications, taking the traditional vocabulary in lower costs and faster way into the current information environment and thus vocabularies play a dual role as organization tool and terminology resources.

Keywords: Semantic wiki; Vocabulary development; Terminology service

(收稿日期: 2014-04-01)