

主题词表的社会应用研究*

周晓英¹, 曾建勋²

(1. 中国人民大学信息资源管理学院, 北京 100872; 2. 中国科学技术信息研究所, 北京 100038)

摘要: 分析主题词表社会应用的现状和目标, 从三个方面说明主题词表社会应用的意义与价值; 在阐释当前主题词表社会应用的难点之后, 提出主题词表社会应用研究的理论建设要求; 并基于国内外主题词表社会应用的实践, 提出针对主题词表社会应用的模式、机制和方案的具体建议。

关键词: 主题词表; 社会应用模式; 应用场景

中图分类号: G254.2

DOI: 10.3772/j.issn.1673—2286.2014.10.001

1 主题词表社会应用的现状分析

主题词表 (Thesaurus) 又称检索词典, 是自然语言中优选出来的语义相关、族性相关的科学术语所组成的一种规范化词典。在文献标引与情报检索过程中, 它是用以将文献、标引人员及用户的自然语言转换为统一的系统语言的一种术语控制工具。

主题词表的社会应用是指将各种类型的主题词表作为实现知识组织、知识检索和知识服务的技术和工具, 将其广泛地应用于社会各种环境中, 实现其应用价值的活动和过程。

主题词表传统上的应用环境主要是在图书馆和信息中心, 图书馆员和信息专家负责主题词表的编制并将编制成果应用在对图书和资料的组织、管理和服务上。由于信息的社会应用程度加剧, 除了图书馆和信息中心之外, 社会的其他环境对于信息的精确组织和管理的要求逐渐增长。我们对国际信息管理实践的调研发现, 当前在发达国家, 主题词表已经被广泛应用在电子政务环境下政府信息的描述和提取中, 电子商务环境的商品管理和大型企业的知识管理领域也在广泛应用主

题词表技术, 医学、教育领域主题词表的应用也较为普遍, 形成了一定的规模。一些国家编制了适应网络需要的电子版和网络版的主题词表, 如农业多语种叙词表AGROVOC经过修订和维护, 演变成了RDF/SKOS-XL资源, 《ERIC网络叙词表》、《人文社会科学电子叙词表》(HASSET)、一体化医学语言系统(UMLS)等, 能适应网络数据库的信息标引和检索的要求; 一些国家将主题词表应用到专门的行业或领域之中发挥作用, 如澳大利亚政府1999年开始应用政府主题词表TAGS和政府职能叙词表AGIFT、新西兰政府使用政府主题词表SONZ和FONZ来描述政府信息和开展政府信息定位服务。

在上述领域无疑都存在一个需要精确定位信息的较大的社会需要。基于上述研究分析, 我们发现, 在需要对信息精确描述和定位的信息利用环境下, 主题词表的利用能够有效改善信息获取环境, 提高知识的过滤和利用效率。

相比于国际先进国家, 我国主题词表在图书情报领域的应用研究较多, 而在除图书情报机构之外的应用还比较落后, 从理论上讲, 目前成果主要是针对词表编

*本研究得到国家自然科学基金项目“面向使用的政府在线信息管理模式与服务规范研究”(编号: 71073168)资助。

制方面的, 其他问题还鲜有研究, 从实践上看, 主题词表的应用不成规模, 缺乏应用规范、缺乏应用机制的保证, 造成了社会普遍性的知识提取和利用困难。因此, 我国目前迫切需要利用主题词表工具来完善知识利用环境, 帮助除图书情报机构之外的各类组织机构或各种知识传播媒介来提高知识的利用效率。

我国主题词表在未来的发展应该要探索主题词表的社会应用渠道, 这样不仅有利于满足当前社会对于知识过滤的强烈需求, 同时也有利于促进主题词表本身的可持续发展, 使其在广泛的社会应用中实现自身的价值。

笔者对主题词表的社会应用目标所持有的观点是: 主题词表的应用场景要尽快从图书馆拓展到社会更加广阔的知识利用环境中, 主题词表的社会化应用应该达到的目标是: 使主题词表成为知识组织的普及性工具, 支持网络环境下的各类型信息应用平台的精确知识描述和知识检索, 提高各类型信息平台信息的可获得性和可用性。

2 主题词表社会应用的意义分析

当前之所以应该重视主题词表的社会应用, 主要基于以下理由:

从词表开发的投入产出角度看: 词表的开发是知识密集型工作, 投入的人力物力财力大。以《汉语主题词表》的第一版编制为例, 1975年的“汉语主题词表编辑组”由全国505个单位1378人参与编写, 还有1048个单位、7519人参加部分的编审工作, 1980年才公开出版。2009年, 中国科学技术信息研究所组织对《汉语主题词表》的修订和重编, 由6所高校和10个科研机构组成的团队参与了该项工作, 这次的编制虽然可以使用更加有效的技术工具, 但仍然是耗费人力时间的工作。网络时代的词表编制虽然编制和维护方式发生了很大的变化, 但其知识密集型、人力耗费、工作量巨大的特征并没有改变。可见, 只有促进词表的社会化应用, 使词表在更广泛的范围普及应用, 才能体现词表编制工作投入的价值。

从词表的社会应用需求看: 学习型组织、知识的加速更新都使知识的检索和提取成为现实社会各类型组织机构的重要工作, 一个完善的知识组织工具必然为组织机构效率的提高产生直接的影响; 组织机构和个人面对的要处理的信息量巨大, 而需要精确定位和提取信息的情形逐渐增多, 这样的环境必须有合适的知识组织工

具的支持, 而主题词表是其中最有效的工具之一。

从词表本身的可持续发展看: 词表的编制往往不是一蹴而就的, 持续的更新是保证其价值的重要因素。开展词表的社会化应用, 实现词表有条件的共享、最大限度的重用, 才能够有利于吸引社会的力量来保证词表的维护和更新, 才能有可靠的机制保证词表成为一个动态更新的、生长中的有机体。此外, 网络环境下的主题词表的应用是词表发展的新机遇, 联合国粮农组织数据库使用的农业叙词表AGROVOC、美国国家医学图书馆的医学主题词表MeSH、美国国会图书馆标题表LCSH等世界著名词表都进行了适应网络应用需要的修订和维护, 词表在网络时代有条件发挥新的更大作用。

综上, 无论从词表本身的投入产出价值分析、从词表自身发展和完善的需要, 还是从满足社会对知识组织工具的期待的角度, 主题词表的社会应用都是当前迫切需要解决的问题, 需要从理论上探索主题词表社会应用的模式, 实践上建立主题词表社会应用的良好机制, 搭建社会应用的平台。

3 当前主题词表社会应用的难点分析

1959年, 美国编制了世界首部叙词表《技术叙词表》, 开启了叙词表编制的时代。1960年美国武装部队技术情报局(ASTIA)编制了计算机使用的主题词表《ASTIA词表》, 据称, 累计到现在, 国外已经编辑出版了上千种叙词表来用作信息检索语言。我国1971年编制出版第一部专业叙词表《航空科技资料主题表》, 1980年出版第一部大型综合性叙词表《汉语主题词表》, 70年代至90年代末期, 已经编制了多达150多部专业的和综合叙词表。尽管其中有些词表在图书馆和信息机构的确发挥着很重要的作用, 但是也有大部分词表没有持续的维护、更新和完善, 有学者曾经统计过我国叙词表修订方面的情况, 发现153部叙词表中仅有22部维持着后续的修订, 其中词表编制了电子版和网络版的也比较少。此外, 走出图书馆和信息机构, 叙词表的社会应用情况更是不容乐观。图书情报人员仍然还在努力开发和完善的词表, 没有得到有效的社会利用, 其社会价值和经济价值没有能够有效挖掘与实现。

3.1 主题词表社会应用认识的缺失

我国学者目前对主题词表的研究成果, 绝大部分

是关于主题词表的编制和建设方面的,对主题词表的应用方面,侧重关于标引人员如何利用主题词表开展标引的研究或词表在图书情报机构中的应用。可以说,目前我国学者对于词表的研究大部分是关于词表本身的研究,解决的都是关于词表如何建设的技术问题、词表之间的互操作问题等等,而对于主题词表“如何与社会产生交互作用?如何提供社会应用和社会服务?词表的分享和再利用该如何做?”这个角度的研究几乎很少学者问津。可以说在主题词表社会应用的认识方面有缺失。

认识缺失究其原因主要是由于:从本源上看,词表的开发是来源于图书馆信息机构的工作需要,词表也是面向图书馆信息机构的标引人员的。图书情报领域的不少人对于词表的看法难以跳出这样的思维定势来考察词表的应用问题。虽然有些有识之士已认识到这个问题,但是由于主题词表新的需求环境与图书馆环境差异很大,需要深入实践、结合业务工作,在实践积累的基础上重新探索主题词表应用问题,这个认识形成过程需要跨学科跨领域的思维和体验,难度比较大。

3.2 主题词表社会应用模式的缺乏

首先,由于网络环境带来信息管理无处不在的变化、带来用户信息获取方式的变化,管理信息的人不再只是图书情报工作者,因此主题词表应用场景产生了极大变化,应用的场景变化后,主题词表新的应用模式尚未建立起来。

其次,主题词表新的应用场景有新的社会需求,而这种需求与图书馆信息中心环境有很大不同,与应用者的业务环境有更多的关联,它需要主题词表应用者是懂业务、懂信息管理、懂信息技术等其他跨学科的相关知识的复合型人才。

第三,新的应用场景下,主题词表的利用从图书情报专业人士转到其他普通从事信息管理工作人员的变化,由技术人员和业务人员提出的知识查询和利用解决方案中,由于他们知识背景的原因,往往没有充分认识和利用主题词表作为知识组织工具的功能,他们之中有些人在不自觉中应用着主题词表的相关技术,但却不了解现有的、已编制的主题词表,因此也不能高质量、有效地采纳已有的词表来节省经费时间和提高效率。

总之,目前主题词表社会应用的困难在于应用模式的缺乏、跨越学科专业造成的信息组织专业人士的认

识不足以及社会认识不足等方面。

4 主题词表社会应用需解决的理论问题

大型综合性主题词表涉及的领域非常广泛,因此其预期可应用领域也非常广,在确定特定的应用场景和应用需求后,建立主题词表与应用业务相结合的应用环境,便能支持业务的知识组织与利用的需要,同时,也能为主题词表本身的修改、更新、维护提供良好的生长环境。在当前中国的环境下,为促进主题词表的社会应用,我们在理论上需要解决的问题如下:

4.1 主题词表社会化应用的需求分析

主题词表的社会应用与我们熟悉的在图书情报机构的应用有较大的不同,理论上应该根据国内外实践梳理出其社会化应用的可能场景,明确其可能有的应用条件和应用机会;此外,还需要明确各种应用场景中的不同应用需求,才能进一步寻求主题词表社会应用的可能性。

当前,主题词表的应用场景已经从纸面转移到了网络上,会以网站应用、平台应用、系统应用和嵌入式应用的应用场景出现。国际先进国家在某些领域已经很好适应了这种转换,词表作为知识组织、知识过滤和知识发现的工具,在众多领域新的应用场景中发挥作用,而我国目前还未能完全实现这种转换。由于实际部门有强大需求,各领域的实践工作者可能使用不规范的方式自行建立了多种多样具有词表和词库功能的工具,花费了人力物力财力,但由于这些词表专业化程度不高、技术水平上尚有欠缺,而实践工作者们并不知道哪里存在或者可以获得那些直接能够满足他们需求的词汇和术语服务。

4.2 主题词表社会化应用的场景构筑

网络环境下主题词表新的应用场景存在以下特点:

1) 主题词表的应用比原来的专业应用环境要复杂得多; 2) 需要适应不同专业背景知识的人员的简单化利用特点; 3) 词表的更新速度加快; 4) 词表合作开发和共同维护的情形增多。主题词表的不同应用场景应有不同的需求,我们应该从理论上深入研究主题词表的不同应用场景以及相应的对主题词表的应用需求,为

推动主题词表在实践中的社会应用打下基础。

4.3 主题词表社会化应用的模式研究

主题词表社会化应用的形式和方式是考察其社会化应用的重要视角,应研究当前国内外主题词表在图书情报机构应用的理论与实践,特别要分类型重点研究主题词表在电子政务、电子商务、医疗健康、教育等环境下的应用实践和相关理论,在此基础上,根据我国的国情和主题词表本身的特征,分析其在我国的应用环境,探讨其社会应用的可行性和可能的应用模式,解决主题词表在图书情报机构之外的社会化应用模式的梳理提炼。

4.4 主题词表社会应用的运行机制

主题词表社会利用运行机制和运作模式是保证主题词表能够有效应用的关键,应分析和总结国内外最佳实践,分析其运行的机理和运作方式,作为我国主题词表应用推广的实践基础,根据主题词表的社会应用实践以及理性分析推理其社会应用的运作方面的经验和规律。考察制度体制建设、规范建设、内容资源建设、用户资源建设等方面的动态开发和更新维护的机制。

5 主题词表的社会应用方案建议

5.1 主题词表社会应用模式

主题词表的编制和修订花费了大量人力物力,理应建立某种机制让其产生社会效益或经济效益,服务于社会需要和服务于经济发展。国际上有成功的经验,值得我们借鉴。

主题词表的社会应用模式按照应用对象可以分为两种:

一是公益性社会应用。公益性社会应用适合于社会公共部门领域的利用,比如政府部门、医疗卫生、社会养老等领域,其特点是将主题词表通过授权免费提供给公共服务的社会机构利用,以降低社会公共服务的成本,提升公共服务的质量。主题词表的建设一般属于信息基础设施建设中的内容,词表是标准化和共享的基础,很多词表的开发和维护是由国家提供建设项目资金和人力保障的,这类词表的基本利用理应是公益性质的。

二是商业性社会应用。商业性社会应用适合于公司

或企业以及其他的专门化利用,可适当对主题词表作针对性的二次开发以适应公司企业以及业务的特殊需要。主题词表虽然是公益性开发和利用的社会资源,但是在词表基础上作专门化的深度开发和专门化的服务是可以适当收取费用的。也可以用这部分的费用贴补公益性利用的支出。

主题词表的社会应用模式按照维护状况也可以分为两种:

一是授权使用或出售之后负责维护。指《汉表》产权所有者的管理机构帮助使用者继续维护词表的使用。

二是一次性出售,不负责维护。指《汉表》将客户需要词表部分出售之后,不再负责维护的情况。

5.2 主题词表的社会应用机制

由于社会实际上存在着对词表、术语的强大需求,在内容管理、信息管理、知识管理、服务管理等领域开展主题词表的社会利用活动是切实可行的。

根据我们对实践成功案例的分析,促进主题词表的社会利用可采取以下机制:

公共领域免费试用促进机制:采用类似数据库产品试用的办法,选择政府和医疗健康领域的网站或系统或平台,通过提供免费词表服务,促进词表的社会影响力的提升,促进词表作用的社会认知度和识别度的提高。相关的实例有:加拿大政府为主题词表设定了一个专门的网站,提供多版本主题词表便利的下载方式,不仅免费提供丰富的检索词入口和精确的索引,还提供实时的在线反馈机制,使得用户体验效果好。

商业合作机制:选择网站或平台或系统开发运行维护公司,采用与公司合作的方式,利用该公司在网站、平台或系统建设和维护过程中与客户的沟通,推进引进网站、平台或系统中对主题词表的利用模块。

合作开发和利用机制:开展国际或者国内的合作,与相关机构在已有主题词表的基础上合作开发专用词表,在合理的范围内共建共享词表成果。在这方面我们可以借鉴核科学词表《INIS/ETDE主题词表》的国际合作机制。该词表由美国国会图书馆组织,由世界范围内多国联合共同开发利用、共享开发成果。《INIS/ETDE主题词表》由国际核信息系统INIS(IAEA的下属部门)与各成员国合作开发,先后开发了具有导航功能的INIS/ETDE多语种主题词表,其中包括完整的词库结构,语种涵盖了国际原子能机构的所有官方语言,

包括阿拉伯文、中文、英文、法文、俄文、西班牙文、德语和日语。

动态更新机制:词表的建立不是一劳永逸的,需要有动态更新机制,可以形成多个版本的分类,包括不经更新的打印版,每年更新的网络版,每年更新一次的PDF版,根据需要随时更新的通告海报(poster)。比如,国外很多政府针对电子政务主题词表专门建立一个网站对其进行专业的在线管理,在网站中不仅提供主题词的索引,同时也设置了反馈评论的功能,允许用户对主题词表进行评论和建议,有些网站还提供了专门的论坛,从而让用户也能参与在线的更新维护,保持词表的可用性,充分适应电子政务发展的需要。如英国政府集成的公共部门词表(Integrated Public Sector Vocabulary, IPSV)所采用的更新机制很值得借鉴。英国电子政府元数据标准eGMS规定采用IPSV来描述政府信息的主题,而IPSV通过建立的专门网站,进行定期的在线更新维护,同时针对每个术语提供comments功能,允许用户自由评论,并且提供IPSV专门的论坛,为更多的用户评论、建议提供平台,从而可以吸纳用户意见和建议,对IPSV作出更好的改进或修订,实现持续的更新维护。

工具化嵌入机制:主题词表作为一种信息产物和知识产品,利用它的优势,将主题词表工具化,嵌入到信息生成和信息检索环境或者系统平台中,优化信息内容的呈现和利用方式,必然大大提高信息处理和利用的效率。

标准化推广机制:主题词表的标准很多,对于其建立、词表构成、发展和利用都有相应的标准,也有学者研究了自1967年至2013年国际标准化组织ISO、美国国家信息标准化组织NISO、互联网联盟W3C、欧洲新闻中心EJC等机构发布的相关标准。本文所强调的标准化推广机制是指在行业和领域推行词表应用标准,比如

在政府信息公开的政务词表的使用、在医疗健康网站的信息质量控制上推行国家标准的制定和实施。

参考文献

- [1] 中华人民共和国国家标准.文献叙词标引规则GB/T3860-95 [S].中国标准出版社,1995.
- [2] AIMS Agricultural Information Management Standards [EB/OL]. [2014-10-06]. <http://aims.fao.org/standards/agrovoc/about>.
- [3] Australia Government. Thesaurus of Australian Government Subjects [EB/OL]. [2014-10-06]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.177.2035&rep=rep1&type=pdf>.
- [4] National Archives of Australia. A standard framework for describing the functions of government [EB/OL]. [2014-10-06]. <http://www.naa.gov.au/records-management/publications/agift.aspx>.
- [5] Functions of New Zealand Thesaurus [EB/OL]. [2014-10-06]. <http://www.vocabularyserver.com/fonz/index.php>.
- [6] Subjects of New Zealand Thesaurus [EB/OL]. [2014-10-06]. <http://www.vocabularyserver.com/sonz/>.
- [7] 贺德方.《汉语主题词表》的回顾与展望[J].情报理论与实践,2010,33(2):1-4.
- [8] 曾建勋,常春.网络时代叙词表的编制与应用[J].情报理论与实践,2009,53(8):8-11.
- [9] 宋翠云.中外叙词表的发展[J].河南图书馆学刊,1995(4):27-28.
- [10] 鲍秀林,吴雯娜.40年来中文叙词表修订情况概览[J].图书情报工作,2013(2):109-113.
- [11] 康艳,张虹,侯汉清.情报检索语言不是“明日黄花”[J].图书情报工作,2007(10):139-142.
- [12] ISQ Information Standards Quarterly [EB/OL]. [2014-10-06]. http://www.niso.org/apps/group_public/download.php/8289/SP_clarke_zeng_isqv24n01.pdf.

作者简介

周晓英,女,1963年生,教授,博士生导师,研究方向:信息组织与信息构建、信息资源管理, E-mail: xyz-ruc@qq.com。
曾建勋,男,1965年生,研究馆员,博士生导师。

Study on the Social Application of Thesaurus

ZHOU XiaoYing¹, ZENG JianXun²

(1. School of Information Resource Management, Renmin University, Beijing 100872, China;

2. Institute of Scientific and Technical information of China, Beijing 100038, China)

Abstract: This paper analyzes the status quo and object of thesaurus in social applications, explains the significance and the value of social applications of thesaurus from three aspects. Based on the analysis of difficulties of social applications of thesaurus, it provides the requirement of theory construction of thesaurus in social applications. With the analysis of the practice of thesaurus social application both at home and abroad, it puts forward the specific recommendations and solutions of social application mode, mechanism and scheme of thesaurus.

Keywords: Thesaurus; Social application mode; Application scenario

(收稿日期: 2014-09-22)