

主题词表在政府信息管理中的应用*

李秀华¹, 周晓英¹, 张萍^{1,2}

(1. 中国人民大学信息资源管理学院, 北京 100872; 2. 中国人民大学图书馆, 北京 100872)

摘要: 采用文献调研和网络调研的方法, 选取英国、加拿大和澳大利亚的应用案例, 通过研究国外主题词表在政府信息管理的应用特点和经验, 结合我国主题词表在政府信息管理中的应用现状, 为我国主题词表在政府信息管理中的应用提供建议, 旨在推动我国主题词表在政府信息管理中的应用和发展。

关键词: 主题词表; 政府信息管理; 应用

中图分类号: G254.2

DOI: 10.3772/j.issn.1673—2286.2014.10.002

主题词表是指按字顺排列、并以参照系统揭示相互关系的概念标识词表, 是将文献内容出现的自然语言规范为检索语言的一种控制术语的工具^[1]。作为规范表达内容一致性的特定序列的词汇语言体系, 主题词表对信息分类和信息检索具有独特的作用, 有助于提高信息的查准率和查全率, 强化信息的可控性和检索的实用性。尤其是在海量信息剧增的网络时代, 主题词表工具有机集成到信息组织和信息检索中, 能够优化信息的呈现, 提高信息利用率。

主题词表在英国、加拿大、澳大利亚等国的政府信息管理中得到了很好的应用, 这些国家不仅编制了专门的政府主题词表, 还将这些词表应用于政府信息管理和政府网站的建设之中, 极大地促进了政府信息展示和检索。相比之下, 我国主题词表在政府信息管理中的应用亟待加强。

本文基于文献调研和网络调研的方法, 选取英国、加拿大和澳大利亚的应用案例, 通过研究国外主题词表在政府信息管理的应用特点和经验, 结合我国主题词表在政府信息管理中的应用现状, 为我国主题词表在政府信息管理中的应用提供建议, 旨在推动我国主

题词表在政府信息管理中的应用和发展。

1 文献综述和研究现状

国内关于政府主题词表的研究集中出现于2003年, 研究重点主要集中于主题词表的编制和构建^[2-6], 主题标引词表在系统应用中的算法研究^[7], 主题词表用于本体的构建等领域^[8]。国外电子政务主题词表研究侧重于词表的编制、实现和功能分析, 其研究主要是基于个体词表基础之上, 回答了什么是电子政务主题词表、怎样构建电子政务主题词表的问题^[9-11]。

综合国内外文献可以发现, 关于主题词表的文献偏重于理论研究和微观层面的应用技术, 缺乏高屋建瓴的对主题词表在政府信息管理中应用功能的研究, 在全文检索技术不断发展的挑战下, 主题词表在政府信息管理中的作用和价值趋于弱化。这个问题值得我们思考和研究。

从国内外应用实践来看, 主题词表在英国、加拿大、澳大利亚等国的政府信息管理和政府网站建设中发展和应用比较成熟, 其成功经验值得我们参考借鉴。

*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”(编号: 71073168)资助。

2 国外主题词表在英国、加拿大、澳大利亚政府信息管理中的应用

2.1 IPSV在英国综合公共部门政府信息管理中的应用

公共部门主题词表(The Integrated Public Sector Vocabulary, IPSV)^[12]是英国公共部门采用的主题词表,用于英国电子政务元数据标准(E-Government Metadata Standard, e-GMS)^[13]中主题项的编码,为e-GMS元数据主题项提供标准的术语,旨在促进英国政府信息的标准化管理,提高政府信息的规范化和工作效率。

IPSV由OPDM(Office of the Deputy Prime Minister)^[14]和eGU(Cabinet Office e-Government Unit)^[15]两个机构基于早期的三个主题词表GCL、LGCL(Local Government Category List)^[16]以及seamlessUK taxonomy^[17]基础上开发编制,2006年4月发布的版本2.00共有3080个正式术语和4843个非正式术语,涵盖了面向内部系统和外部公众的大多数主题,因此IPSV具有良好的行业整体性和广泛性。

2012年10月,英国公共部门信息域-元数据标准工作组在e-GMS元数据标准主题项的应用中要求^[18]:公共部门必须使用通用的受控词表作为对内外发布信息的元数据主题元素,而且这些受控词表必须遵循SKOS(Simple Knowledge Organization Systems, SKOS)^[19],从而保证这些标记信息能够被进一步划分和关联。该要求推动了IPSV在英国公共部门的政府信息组织和检索中的广泛应用。

IPSV具有丰富的检索入口词,包含的内容涉及各行各业,面向的对象用户也涵盖了英国各个级别及各个区域的政府机构,适用于指导各级政府根据统一的标准建立政府网站的信息组织和索引机制,让信息能更准确地传播到各行各业的用户和民众。

针对索引机制和词间结构关系,IPSV共提供了16个高级类别,每个高级类别又分别提供了2-5个低级类别,呈树状结构,每个术语都有一个索引ID来唯一标识,如图1所示。每个术语包含了详情(Details)、范围(Scope)、映射(Mappings)、参考(Reference)、格式(Formats)和评论(Comments)六项具体内容,其中详情描述了术语索引ID和名称以及类型及资源地址等信息;映射提供了该术语上位及相关术语,同时展



图1 IPSV词间结构关系

示了与其有映射关系的术语,以及能链接到的相关网页,由此实现词间关系索引;参考则列出该术语相关网页、文档和其他的资源链接,由此也可以获取到其他相关术语,拓展了词间关系。

为了让IPSV得到更广泛的应用,英国建立专门网站^[20],由Esd-toolkit^[21]开发管理运营,对IPSV的发展历史、词表概述、详细介绍、使用目的、可用版本和格式、如何应用进行标引、自动分类、应用软件及相关法律等内容提供了非常详细的指南。同时,网站提供了IPSV的多个版本和格式(如XML、CSV、WORD和PDF等),各组织及民众无需e-GMS授权即可下载使用。

2.2 CST在加拿大政府信息管理中的应用

加拿大政府信息化程度很高,2001年就提出了政府信息管理的元数据框架,而且采用了加拿大主题词表(Canadian Subject Headings, CSH)^[22]和加拿大政府核心主题词表(The Government of Canada Core Subject Thesaurus, CST)^[23]。CST发展于2004年1月,是目前加拿大政府使用最广泛、影响力度最大的一个词表。

作为加拿大政府信息资源双语专业术语表,CST

有英语和法语两个版本, 其中英文版本有4801个术语, 法语版本有4868个术语, 这两个版本处于平等地位, 同时编撰, 在范畴覆盖和主题词数目上可以互换。CST涵盖了描述加拿大政府以各种形式生成和发布的所有主题信息的词语。CST在国际上率先实现了网络版应用, 并建立了每一个检索词与政府出版物目录库 (Government of Canada Publications Catalogue)^[24]、政府网上信息库 (Canadian Government Information on the Internet Catalogue) 和网站搜索引擎 (Canadian Site Search engine) 的链接查询。

为了促进CST的专业推广和应用, 加拿大政府建立了一个网站 (www.thesaurus.gc.ca) 以提供专门服务。该网站提供族首词索引和字顺索引两种形式链接, 通过“Specificity”、“Exhaustivity”、“Consistency”和“Indexing language”四个索引规则来达到精确的索引, 并显示关联词和词间关系。同时还提供搜索引擎形式查询, 可按单词切分和字符串进行检索。

通过词表的精确机制、丰富检索词入口及词间的强关联性, 既可以增加自然语言和政府术语的转换, 又可以充分利用主题词表参照系统特点, 建立多种关联关系, 从而提高公众和不同部门政务用户检索的查全率、查准率和用户的可用性、可操作性等, 使词表能够更容易得到拓展。例如浏览路径“ET Education and Training→Colleges”时, 可以找到“Colleges”术语内容体系, 如图2所示。

Colleges		Show Hierarchy
French	Collège	
Subject Category	ET Education and Training	
Used For	CAAT CEGEPs Colleges of Applied Arts and Technology Community colleges	
Broader Term	Educational institutions	
Narrower Term	Military colleges	
Related Term	Postsecondary education Universities	

图2 CST术语“Colleges”内容体系

CST用于网站的词表浏览和搜索, 帮助联邦政府各部门和机构制定受控主题词汇的标准。按照要求, 所有加拿大政府网站必须采用题目、来源、语种、日期和受控主题五项基本数据项作为网络资源的元数据标准描

述网络资源。其中受控主题的标引就来源于CST, 并定期在网上发布政府各部门提供的补充词。

2.3 AGIFT在澳大利亚政府信息管理中的应用

澳大利亚应用于政府网站的建设和电子政务的服务中的主题词表主要有: 澳大利亚政府主题词表 (Thesaurus of Australian Government Subjects, TAGS)^[25]和澳大利亚政府交互职能主题词表 (Australian Governments' Interactive Functions Thesaurus, AGIFT)^[26]。

TAGS是澳大利亚信息管理办公室 (AGIMO) 主持编制的供联邦机构使用的高级主题词表, 遵循澳大利亚定位服务 (Australian Government Locator Service, AGLS)^[27], 基于DC元数据标准, 用于政府机构在线资源发现元数据标准中描述的主题元素 (Subject)。

AGIFT则由澳大利亚国家图书馆发布, 用以完善AGLS职能元素 (Function), 帮助终端用户在检索信息时将自然语言与官方表达的转换, 帮助用户清除检索语言障碍, 促进政府信息的检索和发现。TAGS对应于AGLS元数据的主题项, AGIFT对应于AGLS的职能项, 下面对AGIFT进行重点应用分析。

AGIFT是一个描述涵盖澳大利亚公共组织、州和地方政府职能的的综合的、高度概括的主题词表, 主要是为了描述网络基础下政府信息和服务的在线资源, 旨在帮助终端用户在政府网站上用自然语言进行精确而方便地检索政务术语, 例如, 检索“School”或者“Kindergarten”等相关词语时, 可以匹配到政府部门使用的词语“early children education”。AGIFT包含了25个大类, 每个大类包含了2级和3级、非正式和相关联的术语, 同时它还提供范畴注释来描述每个正式术语所涵盖的含义及其参考。

AGIFT早在1999年就应用于澳大利亚联邦政府网站 (australia.gov.au), 为电子政务提供如下功能:

①为澳大利亚政府机构使用澳大利亚政府定位服务 (Australian Government Locator Service, AGLS) 元数据标准中的职能项提供标准术语。②当用户不确定使用哪个术语或者不清楚哪级政府负责自己要求的信息服务时, AGIFT能帮助用户搜索政府入口点。③为澳大利亚政府各级机构根据分类的需求开发基于机构的

功能词表提供框架。

由此可见, AGIFT作为基本词语元素, 为政府管理网站信息管理提供了选词材料, 尤其是在建立有效的检索目录时, 对原有的术语进行整合, 根据用户检索习惯和检索需求, 提取部分有效的术语, 提高网站信息资源管理水平。

3 主题词表应用的成功因素分析

3.1 接轨国际标准, 重视元数据应用

综上述举例可知, 那些得到成功应用的电子政务主题词表在开发和编制之初就参考引用了相关国际标准, 例如ANSI Z39.19-198、ISO 5964: 1985、ISO 2788-1986和ISO 25964-1:2011等, 这些标准对于主题词表的编制具有重要的指导意义。同时, 电子政务主题词表主要应用于电子政务的元数据标准, 将主题词表纳入到元数据标准, 充当职能项(function)或者主题项(subject)等, 主题词表的应用与元数据标准息息相关。利用元数据标准, 充分参考国际标准, 不仅能使词表的编制和发展更具指导性, 还能使主题词表的应用更加广泛和有效。

3.2 丰富检索词入口, 词表广泛, 适合各级政务需要

主题词表在政府信息管理中的重要应用就是改善政府网站的信息搜索功能和效果。编制良好的主题词表有着丰富的检索词入口, 包括大量的正式术语和非正式术语, 提供多种索引方法, 包括全文搜索、字符串搜索、字母顺序索引等。此外, 主题词表提供术语的上位、下位及其相关的术语等关联信息, 极大地增强了词与词之间的关系和联系。

能够在政府信息管理中广泛应用的主题词表, 如IPSV和CST, 其规模都比较大, 包含的术语数量众多, 基本涵盖了通用的与政府相关的术语, 能够适应各级政府的政务需要, 避免重复编制各级政府主题词表, 降低主题词表的编制和使用成本, 同时使政府信息管理更加规范和一致, 提高政府网站的使用效率。

正是基于此点认识, 加拿大政府不惜投入大量人力物力支持发展CST, 并应用于政府网站的建设。也正是基于这种考虑, AGIFT被开发并被应用于澳大利亚

政府网站, AGIFT为AGLS的职能项提供了标准的术语, 与网上在线主题词表的30万条自然术语进行了映射, 让民众在使用政府网站时通过自然语言就能搜索到准确、全面的政府信息。

3.3 专业稳定的更新维护与有效的在线管理

主题词表的更新维护与在线管理对网络环境下的主题词表的应用具有重要影响。专业稳定的更新维护与有效的在线管理, 能保证主题词表应用的长久生命力和活力, 从而增强其应用价值和社会价值。

CST的更新维护主要由加拿大图书馆和档案馆负责, 因其拥有专业的信息管理背景, 能保证主题词表更新与维护的专业规范性。词表的在线管理则由加拿大环境部等机构负责, 提供该词表在网络应用方面的支持。同时, 网站也提供反馈机制, 让用户能实时反馈对主题词表的建议与意见, 使主题词表更具广泛性和代表性, 既能在一定程度上反映用户的实际需求, 也能保证词表的专业性, 并指导政府网站的应用。

3.4 综合利用现有词表以降低成本

国外政府在开发编制词表的时候, 会考虑是否能综合利用已有的主题词表, 例如英国IPSV就是基于GCL、LGCL和seamlessUK taxonomy三个已有主题词表成功编制出来的, 降低了开发成本, 更重要的是很多政府网站在应用IPSV时可以比较容易地进行移植和拓展应用, 这样既能降低主题词表在网站应用时的难度, 也能从整体上降低其应用的成本。

3.5 政府主导下的专业机构合作

英国IPSV的编制和开发基于22个国家级电子政务项目, 由ODPM和GU两个机构合力开发编制, 同时也获得了本地电子政务标准团体(Local e-Government Standards Body, LeGSB)和Essex郡议会(Essex County Council)的支持, 机构间分工明确, 密切合作, 共同推动IPSV的发展和应用。

CST之所以能得到广泛的应用, 也离不开多方机构合作和先进的管理支持。最初以政府相关部门为主导研制CST, 随后由图书馆和档案馆进行专业管理和维

护, 环境部对CST网站进行经营管理, 各级政府网站的应用, 以及诸多大学等相关机构参与合作, 使得该主题词表从开发到应用和推广都有着良好的合作基础, 让主题词表变得更具权威性和社会性, 从而保证了CST的持续发展和应用。

3.6 有效的应用推广方式

IPSV通过网站详细介绍了其基本情况和使用方法, 提供了在线查询和多版本多格式免费下载, 并设置专门的论坛, 使用户在方便地使用、共享IPSV的同时, 能对IPSV进行评论, 提出自己的建议和看法。更重要的是, 用户使用IPSV不需要获得e-GMS的授权, 利用IPSV的自由度高, 能吸引更多的用户参与应用IPSV, 这些都是推广IPSV应用非常有效的方式, 值得我国政府在推广主题词表应用时借鉴。

主题词表的推广应用与相关工具的开发应用密不可分, 例如针对IPSV, 诸多商家开发出各种应用软件, 如Semaphore^[28]、Searchlight^[29]、conceptSearching^[30]、EasySite CMS^[31]、iCM^[32]等, 这些系统具有利用IPSV进行分类、支持所有e-GMS元素的标引等应用广泛的功能, 极大地促进了IPSV的推广和应用。

4 结语

我国政府信息管理工作也应用了一些政府主题词表, 主要体现在: 一是中央机构及各个地方政府颁布的公文主题词表, 例如《国务院公文主题词表》^[33]、《江苏省人民政府主题词表》^[34]、《内蒙古自治区人民政府公文主题词表》^[35]; 二是各个政府网站在信息公开上采用主题分类的方式建立的信息资源目录及交换体系, 例如《国务院办公厅政府信息公开目录》^[36]、《广东省政府信息公开目录系统》^[37]; 三是科研团体在承担国家重点工程项目时的研究成果, 例如赵新力主编的《综合电子政务主题词表字顺表》。

与国外的政府主题词表应用相比, 国内的政府主题词表的应用和影响范围非常有限, 词表影响的主要是颁布机构直接主管的部门和机构, 缺少与普通民众之间的交互, 难以为企业或者个人提供服务, 同时缺乏专业机构进行维护和更新, 词表的推广和应用比较困难。

因此, 我们需要针对国内主题词表的应用现状, 在

主题词表的开发和编制、维护和发展, 以及主题词表的推广和应用等方面, 积极参考借鉴英国、加拿大和澳大利亚等国政府主题词表的成功应用, 推动我国主题词表在政府信息管理中的应用和发展。

参考文献

- [1] 袁世全, 冯涛. 中国百科大辞典[M]. 北京: 华夏出版社, 1990: 444.
- [2] 倪静. 电子政务主题词表的编制及应用研究[D]. 北京: 中国科学技术信息研究所, 2003.
- [3] 倪静, 赵新力, 钱起霖. 国外电子政务主题词表编制及网络应用的比较分析[J]. 情报学报, 2003, 22(5): 565-571.
- [4] 盛苏平, 刘春燕, 赵新力, 等. 《电子政务主题词表》编制及应用系统开发[J]. 中国信息导报, 2006(3): 37-39.
- [5] 仲云云. 电子政务主题词表的构建及应用研究[D]. 南京: 南京农业大学, 2007.
- [6] 贾君枝, 武晓宇. 基于FAST的综合电子政务主题词表分面式改造[J]. 图书情报工作, 2014(8): 105-110.
- [7] 倪静, 赵新力, 李惠安. 电子政务主题词表应用系统中主题标引算法研究[J]. 高技术通讯, 2003(10): 15-19.
- [8] 卢懿, 张波. 基于政务主题词表的政府信息本体构建及应用[J]. 计算机工程与设计, 2010(3): 521-524.
- [9] CRAIG A, SCHRIAR S. The Find-It! Illinois controlled vocabulary: Improving access to government information through the Jessica subject tree [J]. Government Information Quarterly, 2001, 18(3): 195-208.
- [10] CHEN C-C, YEH J-H, SIE S-H. Government ontology and thesaurus construction: A Taiwanese experience [C]// Proceedings of digital libraries: implementing strategies and sharing experiences, lecture notes in computer science, 2005, 3815: 263-272.
- [11] GAO Wen-fei, ZHAO Xin-Li. Constructing the Relationship between Concepts in Government Ontology Based on E-Government Thesauri [C]// Proceedings of 2008 International Conference on Management Science & Engineering (15TH), 2008, vol(I&II): 116-121.
- [12] IPSV [EB/OL]. [2014-09-15]. <http://doc.esd.org.uk/IPSV/2.00.html>.
- [13] e-GMS [EB/OL]. [2014-09-15]. <http://www.esd.org.uk/standards/egms/>.
- [14] OPDM [EB/OL]. [2014-09-15]. <http://www.dpm.cabinetoffice.gov.uk/>.
- [15] e-GU [EB/OL]. [2014-09-15]. <http://webarchive.nationalarchives.gov.uk/20070603164512/archive.cabinetoffice.gov.uk/e-government/>.

- [16] LGCL [EB/OL]. [2014-09-15]. <http://www.esd.org.uk/standards/lgcl/>.
- [17] seamlessUK taxonomy [EB/OL]. [2014-09-15]. <http://dip.semanticweb.org/documents/D9.3Annex-SeamlessUKtaxonomyXML.pdf>.
- [18] esd standards[EB/OL]. [2014-09-15]. <http://standards.esd.org.uk/?uri=list%2Fsubjects&tab=details>.
- [19] SKOS [EB/OL]. [2014-09-15]. <http://www.w3.org/2004/02/skos/>.
- [20] IPSV [EB/OL]. [2014-09-15]. <http://doc.esd.org.uk/IPSV/2.00.html>.
- [21] esd-toolkit [EB/OL]. [2014-09-15]. <http://www.esd.org.uk/esdtoolkit/default.aspx>.
- [22] CSH [EB/OL]. [2014-09-15]. <http://www.collectionscanada.gc.ca/6/23/index-e.html>.
- [23] CST [EB/OL]. [2014-09-15]. <http://www.thesaurus.gc.ca/default.asp?Lang=En&n=E5807AB0-1>.
- [24] Government of Canada Publications Catalogue [EB/OL]. [2014-09-15]. <http://publications.gc.ca/site/eng/ourCatalogue.html>.
- [25] TAGS [EB/OL]. [2014-09-15]. <http://www.egov.vic.gov.au/websitetractice/taxonomies-thesauri/thesauri-archive.html>.
- [26] AGIFT [EB/OL]. [2014-09-15]. <http://www.naa.gov.au/recordsmanagement/publications/agift.aspx>.
- [27] AGLS [EB/OL]. [2014-09-15]. <http://www.agls.gov.au/>.
- [28] Semaphore [EB/OL]. [2014-09-15]. <http://www.smartlogic.com/landingpages-ipsv>.
- [29] Searchlight [EB/OL]. [2014-09-15]. http://www.cintra.com/webapp/cms/cms00000/index_website.jsp#3040,SW,28.
- [30] conceptSearching [EB/OL]. [2014-09-15]. <http://www.conceptsearching.com/wp/products/concept-searching-suite/conceptsearch/>.
- [31] EasySite CMS [EB/OL]. [2014-09-15]. <http://www.easysitecms.net/>.
- [32] ICM [EB/OL]. [2014-09-15]. <http://www.gossinteractive.com/contentmanagement>.
- [33] 国务院公文主题词表[EB/OL]. [2014-09-15]. <http://www.sndrc.gov.cn/view.jsp?ID=7876>.
- [34] 江苏省人民政府公文主题词表[EB/OL]. [2014-09-15]. <http://www.jnqmqj.gov.cn/law/show.aspx?id=91&cid=34>.
- [35] 内蒙古自治区人民政府公文主题词表[EB/OL]. [2014-09-15]. <http://www.imnc.edu.cn/dpt/news/xbzcfg2006/5-3/114302800.htm>.
- [36] 国务院办公厅政府信息公开目录[EB/OL]. [2014-09-15]. http://www.gov.cn/xxgk/pub/govpublic/jianjie_822.htm.
- [37] 广东省政府信息公开目录体系[EB/OL]. [2014-09-15]. <http://zwgk.gd.gov.cn/006939748/>.

作者简介

李秀华, 女, 1988年生, 硕士, 研究方向: 信息资源管理。

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, 通讯作者, E-mail: xyz-ruc@qq.com。

张萍, 女, 1973年生, 博士生, 研究方向: 信息组织。

Thesaurus Application in Government Information Management

LI XiuHua¹, ZHOU XiaoYing¹, ZHANG Ping^{1,2}

(1. School of Information Resource Management, Renmin University, Beijing 100872, China; 2. Library of Renmin University, Beijing 100872, China)

Abstract: Based on the research methods of literature research and Internet analysis, with the case study of British IPSV, Canadian CST and Australian AGIFT, this paper explores the application characteristics and method of the foreign thesaurus in government information management. Considering current application of thesaurus in Chinese government information management, it provides some advice on the application of thesaurus in Chinese government information management to promote the application and development of the thesaurus in Chinese government information management.

Keywords: Thesaurus; Government information management; Application

(收稿日期: 2014-09-22)