

网络环境下信息组织的 创新与发展

全国第五次情报检索语言发展方向
研讨会论文集

戴维民 赵建华 汪东波 贺德方 主编

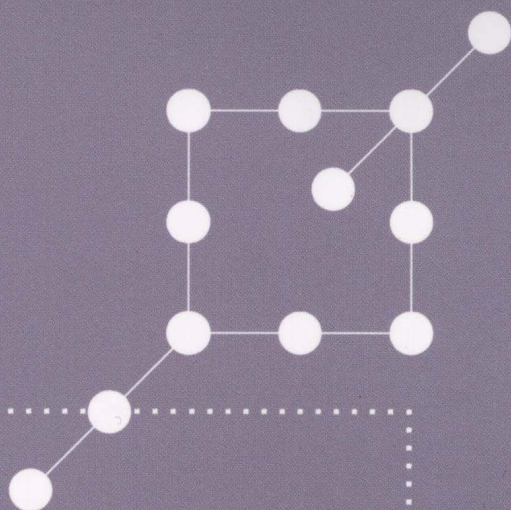


圖 國家圖書館出版社

网络环境下信息组织的创新与发展

——全国第五次情报检索语言发展方向研讨会论文集

戴维民 赵建华 汪东波 贺德方 主编

国家图书馆出版社

图书在版编目(CIP)数据

网络环境下信息组织的创新与发展：全国第五次情报检索语言发展方向研讨会论文集/
戴维民等主编. —北京：国家图书馆出版社，2009. 12

ISBN 978 - 7 - 5013 - 4220 - 4

I. 网… II. 戴… III. 情报检索—检索语言—学术会议—文集 IV. G254.0-53

中国版本图书馆 CIP 数据核字（2009）第 225774 号

书名 网络环境下信息组织的创新与发展——全国第五次情报检索语言发展方向研讨会论文集
著者 戴维民 赵建华 汪东波 贺德方 主编

出版 国家图书馆出版社（原北京图书馆出版社）

（100034 北京市西城区文津街 7 号）

发行 010 - 66139745 66151313 66175620 66126153

66174391（传真） 66126156（门市部）

E-mail btsfxb@nlc.gov.cn（邮购）

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 北京联兴盛业印刷股份有限公司

开本 787 × 1092（毫米） 1/16

印张 25.75

版次 2009 年 12 月第 1 版 2009 年 12 月第 1 次印刷

字数 630（千字）

书号 ISBN 978 - 7 - 5013 - 4220 - 4

定价 72.00 元

目 录

迈向知识组织的新时代

——在“全国第五次情报检索语言发展方向研讨会”上的总结报告 戴维民(1)

我研究情报语言学的若干心得和收获

——自述学术思想 张琪玉(15)

信息组织课程建设与学科发展 戴维民(22)

美国信息构建协会及其对信息构建的贡献和启示 周晓英(31)

分面分类发展路向探析 范 炜(36)

情报检索语言在网络环境中的发展趋势探究 陈俊凤 何晓萍(44)

Internet 环境下情报检索语言的语用研究刍议 朱震远(50)

知识组织系统及其描述与评价初探 薛春香 侯汉清(57)

融于数字图书馆的知识组织系统模型构建探究 姚慧君 徐庆有(65)

基于本体的网络知识组织系统研究 蒋 玲(71)

面向学科数字信息群的知识组织框架研究 敖 毅(75)

中国、美国和英国叙词表编制国家标准比较研究 刘 华 沈玉兰 曾建勋(82)

国外叙词表在网络环境中的应用现状 熊 霞 常 春(90)

电子化环境下的叙词规范控制模式 丁建勤(98)

网络时代专业叙词表选词规则实践与讨论 常 春 吴雯娜(107)

叙词表编制中等同关系获取方法 王世清 吴雯娜 常 春(114)

分类主题一体化中特征词的研究 赵建国等(120)

《汉语主题词表》的回顾与展望 贺德方(124)

OntoThesaurus 与 SKOS 的比较研究

——以及对建立中文 KOS 形式化表示标准体系的建议 曾新红(130)

《中国分类主题词表》网络化研究:从 MARC 到 SKOS 喻 菲(139)

《中国分类主题词表》的 SKOS 描述自动转换研究 刘丽斌等(148)

基于《中分表》的国家数字图书馆知识组织思考 卜书庆(155)

汉语叙词表主题标引的瓶颈与发展方向 曹玉强(161)

我军信息检索语言编制管理工作发展历史、现状及今后展望 赵建华(166)

台湾地区文献分类法厘定特色探析 黄剑年(171)

网络环境下《中图法》第 5 版修订发展趋向的探讨 宋剑祥(178)

面向行业的网络信息组织 郑 敏(185)

基于 Lucene 的企业搜索引擎设计与实现 李冰鹏(191)

我国公文主题词表分析研究及其发展探讨 郝嘉树(198)

传统信息组织理论在信息服务中的应用实例

——《号百信息分类法》编制技术初探 李 华 康 艳(208)

中国地方志物产分类方法研究

——以明、清、民国广东方志物产为例 衡中青(215)

主题图在特色数据库构建中的应用 刘 丹 羌 丽(224)

MeSH 主题词在 PubMed 数据库检索中的作用研究 邱君瑞 陆 敏(234)

电子文件管理元数据标准分类方案的比较研究 张 茜(239)

论电子文件管理元数据元素语义的本体描述 张正强(248)

主题聚类方法的探索与实践 章成志(256)

词频与自动标引之关系研究 傅 亮(269)

事实知识抽取中的语义自足性问题研究 张运良 朱礼军 乔晓东(280)

基于《中图法》的多层自动分类影响因素分析 何 琳 刘 竞 侯汉清(285)

基于概念关系梳理的多领域科技主题词表集成研究 张 宇(294)

基于情报检索语言互操作的集成词库构建

——以教育集成词库为例 刘华梅(304)

主题检索系统的优化技术研究 徐 震(311)

基于机构知识组织的历史文献检索方法 李 超(318)

情报检索语言与自动标引网络信息资源初探 王爱华(324)

中文图书在版编目数据中分类标引的质量评析 贾宇群(330)

基于 XML 的 MARC 信息组织实现及其相关技术研究 王兰成(336)

基于载体 MARC 特征信息的自动分类研究 王兰成 陈 洋 刘晓亮(343)

中国当代索引学的精髓

——张琪玉教授的索引学思想研究 曹树金 姚 瑶(350)

现代的索引就是知识数据库 祝方林(357)

Web2.0 环境下的知识搜索引擎发展研究 彭 陶 马张华(361)

Web2.0 环境下的个性化信息检索服务 周胜利(370)

Folksonomy 检索初探 韩建新(375)

标注行为、标签及其组织模式的优化

——以个人学术空间为例 包冬梅(381)

大众标注的性能及其在图书馆中的应用 倪 皓(390)

大众分类与语义网 王玉媛(395)

博客信息自组织模式研究 罗 昊(401)

迈向知识组织的新时代

——在“全国第五次情报检索语言发展方向研讨会”上的总结报告

戴维民(南京政治学院上海分院)

由国家图书馆《中图法》编委会、全国信息与文献标准化技术委员会、全军军用主题词表编管会和南京政治学院上海分院联合举办,高等教育出版社协办,南京政治学院上海分院承办的“全国第五次情报检索语言发展方向研讨会”于2009年6月11—13日在上海召开。经过与会代表的共同努力,会议取得了丰硕的成果。本次会议可谓内容丰富,交流热烈,是一次开得非常成功、非常专业的学术会议。下面,受会议各举办单位的委托,由我对这次会议作个学术总结,因为准备时间匆忙和我本人水平的限制,以下总结难免有不当之处,请各位代表见谅。

1 会议的主要特点

本次会议的特点可以概括为以下几点:

(1) 主题鲜明,重点突出

本次会议以“网络环境下信息组织的创新与发展”为研讨主题,在这一主题下,研讨内容广泛而深入,包括情报检索语言在网络环境中的发展与应用,网络环境下检索语言标准研究及其发展,网络环境下信息组织方法与模式研究,数字图书馆知识组织等。提交的论文重点突出:传统信息组织工具的网络化、本体化改造;知识组织相关标准、规范、体系结构的研究;词库集成、互操作;自动标引、自动分类等问题研究;信息组织理论、方法和工具的行业应用;大众分类的优化研究等都是大家讨论的重点。今年又适逢张琪玉教授八十寿辰,大会安排专场报告会研讨了张琪玉先生情报语言学思想,张琪玉教授本人也亲自作了自己三十多年来从事情报语言学教学研究的体会报告。这次会议研究内容的深入给大家印象深刻,信息组织领域的年轻后辈展示的学术潜力更令人鼓舞。

(2) 涉及广泛,研究深入

在本次会议入选论文中,很多问题的研究都较上次会议更为深入,不少论文都有一定的创新性。如:卜书庆的《基于〈中分表〉的国家数字图书馆知识组织思考》、范炜的《分面思想的发展路向分析》、章成志的《主题聚类方法的探索与实践》、曾新红的《Onto Thesaurus 与 SKOS 的比较研究》、李华的《传统信息组织理论在信息服务中的应用实例——〈号百信息分类法〉编制技术初探》、付亮的《词频与自动标引之关系研究》、常春的《网络时代专业叙词表选词规则实践与讨论》等论文都有鲜明的观点和创新性的认识。

(3) 理论先导,实践指向

本次会议论文理论密切联系实际,许多论文的选题、研究都是基于实践的研究、原型系统的构建、结合信息组织和检索的实际需求,很多观点、创新都是经过大量研究和实践的总结,论

述有的放矢、言之有物,令人欣慰。

(4) 人才辈出,后生可畏

本次会议是我国信息组织界人才集聚的盛会,从八十岁的、六七十的老一辈专业功底深、积淀浓厚的元老,到中青年学术骨干,80后的新秀齐聚一堂,共同探讨我们情报检索语言、信息组织、知识组织的发展方向问题。青年教师、博士生占据了论文的一半,可谓人才济济。年青一代如范炜、薛春香、章成志、张运良、付亮、包冬梅、刘华梅、张宇、何琳、李华等的论文让我们感觉到了后生可畏,我们的信息组织领域后继有人。

2 信息组织发展的简要回顾与总结

2.1 20世纪以来世界范围的信息组织发展线索

从图书分类、情报检索语言到信息组织、知识组织,概念的演变预示着一个学科或一个技术领域的递进发展,无论是国内、国外,还是上个世纪、新世纪之初,它所展示的技术路线有着很多相通之处。

我在《20世纪情报语言学》^①长文中系统总结了一个世纪的情报检索语言发展成就和基本线索,展望了21世纪情报检索语言发展方向。

对20世纪情报检索语言产生划时代影响的是1876年同时发生的两件事,一是1876年《杜威十进分类法》第一版出版,开辟了图书分类法的新纪元;二是1876年克特写成了《字典式目录条例》,奠定了标题法的基础。总体上说,20世纪情报检索语言的发展就是分类法和主题法两大主流体系并行发展的结果。此后的一些重要事件,如1933年阮冈纳赞的《冒号分类法》出版,1951年陶伯建立了单元词卡系统,1947—1950年期间穆尔斯创造的叙词法,1958年卢恩创制的题内关键词索引等都是对两大语言体系的推动与发展。20世纪后期,特别是90年代网络技术的快速发展,使自然语言在信息检索中得以广泛应用。

20世纪情报检索语言的发展,根据情报检索语言主流来划分,主要有三个时期:(1)20世纪初至50年代为体系分类检索语言和标题词语言时期。这一时期全世界共编制了几十部大型综合性体系分类法,至少千部以上的专业分类法,数百部标题表。尽管30年代分面分类法就已诞生了,对分类法和所有检索语言理论都产生了重要影响,但对检索语言实践影响还很有限。(2)60年代至90年代初为叙词语言和分析—综合分类检索语言时期,叙词语言成为这一时期的主流,世界各国编制了数千部叙词表,分面分类法易用化研究有了重大突破,开始扩大应用范围。(3)90年代,互联网技术出现以后为自然语言和规范语言相结合使用时期。自然语言检索技术和检索语言网络化技术有了突破性进展,成为研究的热点和应用方向。

2.2 60年来我国信息组织发展的基本经验

新中国建立以后,我国情报检索语言发展进入了新阶段。建国之初,我国广泛吸取了前苏联的图书分类法编制的实践经验,除编制了《山东法》、《东北法》等过渡性的分类法外,还编出了《人大法》、《科图法》、《中小型表草案》、《武大法》以及数量众多的专业分类法。这个时期也注意吸取西方的经验,如在科技情报界广泛采用《国际十进分类法》,但随着“文化大革命”的

^① 戴维民. 20世纪图书馆学情报学. 北京:北京图书馆出版社,2002

开始,很快遭到批判,关上了“朝西”的大门。但1950—1970年这个时期,我国情报检索语言仍取得了辉煌的成绩,最具代表性的是1970年代的《中国图书馆图书分类法》和《汉语主题词表》的编制成功和推广应用,标志着我国的情报检索语言的两大体系的基本建立,为我国情报检索语言的逐步走向统一和标准化奠定了坚实的基础。

进入20世纪80年代,随着我国改革开放的不断深入,我国图书情报事业也得到了巨大的发展,情报检索语言理论与技术得到全面发展。具体表现在:分类检索语言和叙词检索语言技术日趋成熟,并形成了较为完备的情报检索语言体系;情报检索语言应用技术有了重大突破,如情报检索语言计算机编制与管理技术,自动标引和机辅标引技术等都达到应用水平;自然语言应用技术得到开发研究;国外情报检索语言的理论与方法被广泛应用,如分面分析理论技术,分类主题一体化技术等;标准化和兼容化体系基本建立;情报语言学学科体系逐步完善,并成为图书馆学、情报学重要的分支学科。我国情报语言学理论水平和情报检索语言的应用技术达到了国际领先的水平。

进入21世纪,随着网络技术的飞速发展,我国情报语言学理论、方法和技术进一步发展,面向更广,适应网络信息组织需要的信息组织技术与方法得到充分发展,建立了信息组织学科领域。

2005年6月,在“全国第四次情报检索语言发展方向研讨会”上,我提交的《21世纪我国信息组织理论与技术进展》^①梳理了新世纪之初我国信息组织领域的最新成果。1999年6月,在“全国第三次情报检索语言发展方向研讨会”上我提交的长篇报告《中国情报检索语言五十年》^②,比较系统地概述了我国情报检索语言理论与方法技术的成就。1989年,建国40周年之际,张琪玉和丘峰撰写了《我国情报检索语言四十年文献简述》^③,从文献的角度,总结了我国40年情报检索语言领域的主要成果。这些报告和总结勾画了我国信息组织的发展线索,展示了我国信息组织方面的主要成就,揭示了各个发展阶段的时代特点和基本规律,同时也综合反映了我国60年来信息组织领域成果的理论性、应用性、时代性、国际性。

总结60年来发展历程和丰硕成果,我国信息组织领域的基本经验主要有六个方面:

(1)坚持以马列主义毛泽东思想为指导,坚持中国自主的发展模式,走出了一条具有中国特色、时代特征,国际水平的发展道路。

(2)坚持继承和发展相结合,学习借鉴与自主创新相结合,构建了具有中国特色、应用性很强的信息组织工具体系。

(3)坚持理论联系实际,坚持理论先导,创建了具有系统理论体系的分支学科领域,成长了一批具有很高知名度的专家学者和学科带头人,积淀了一批学科代表作。

(4)坚持百花齐放、百家争鸣的学风,通过召开学术会议、笔谈,广泛开展学术争鸣,研究解决了理论和实践中的许多重要课题。

(5)坚持教育为先,人才为本,建立了高层次人才培养体系,保持了学术研究的繁荣和实践的创新发展。

(6)坚持与时俱进,面向世界,着眼未来,积极探索新理论新技术,不断拓展研究和应用领

① 见:网络时代的信息组织. 北京:北京图书馆出版社,2006:1-21

② 见:面向21世纪的情报语言学. 北京:北京图书馆出版社,2000:13-49

③ 见:情报检索语言论文选. 北京:书目文献出版社,1990:1-39

域,保持了可持续的发展态势。

2.3 主题不断演进的五次全国会议

每一次全国情报检索语言发展方向研讨会都对我国情报检索语言的发展起到检阅、总结、导向和推动作用。在我国情报检索语言理论研究方面都具有里程碑意义。

第五次全国会议(2009年,上海)。时逢新中国建国60周年,同时也是为了研讨网络环境下信息组织的新理论、新方法和新技术。会议的主题是“网络环境下信息组织的创新与发展”,在这一主题下,研讨内容广泛而深入。

第四次全国会议(2005年,上海)。第三次会议以来的6年,是情报检索语言发展的一个重要时期,也是该领域理论和实践成果丰硕的阶段。在网络信息序化组织的迫切需求下,“网络时代的信息组织”理所当然地成为会议的主题,研讨的内容涵盖了网络信息组织领域中的主要理论、技术、方法及重大现实应用。从概念、标准体系、本体语言、本体编制工具等宏观的层面全面介绍语义网、本体则是这一时期研究的一大热点。此后的研究,逐渐引向深入,特别是本体如何构建、数字图书馆知识组织系统的构建问题,成为研究的热点。

第三次全国会议(1999年,上海)。从这次会议开始,会议名称规范地改为“全国情报检索语言发展方向研讨会”。会议的主题是“面向21世纪的情报语言学”,会议重点研讨了中国情报检索语言的发展历史、情报检索语言的计算机化和电子化、自然语言情报检索技术、网络环境下情报检索语言、情报语言学系列课程建设和教学法、21世纪情报语言学发展趋势等问题。会议达成“受控语言与自然语言相融合”、“检索语言要适应网络应用必须向易用化方向发展”、“有效的情报检索必定是受控的”等基本共识。此后的情报检索语言的研究重点和热点逐步转向了面向网络信息组织、人工语言与受控语言的结合、如何在自由的网络信息检索中引入受控的因素优化搜索精度等方面。

第二次全国会议(1996年,北京)。本次会议的名称是“全国分类法、叙词表发展学术研讨与成果展示会”,会议提出了情报检索语言面向计算机和网络环境适应的问题。此后,计算机和网络环境下的情报检索语言研究与应用成为一个热点。这次会议还全面展示了我国分类法和主题词表编修成果。

第一次全国会议(1988年,北京)。本次会议名称是“全国汉语叙词表发展方向研讨会”,会议研讨交流了改革开放以来情报检索语言研究的成绩、存在的问题,重点是叙词表的编制。在此次会议之后,我国出现了词表编制或修订的高潮,推出一大批有影响的主题词表。

3 研讨热点评述

全国第五次情报检索语言发展方向研讨会是一次新成果的汇集,反映了新进展、新热点。收入会议论文预印本的有53篇论文。以下部分是择其热点和要点作一概述。

3.1 知识组织相关标准与规范研究

叙词表编制国家标准是对叙词表的编制予以规定并由国家标准机构批准的规范性文件,它对于促进叙词表编制的规范化和词表之间的数据交换和共享有着重要作用。随着美国、英国、ISO有关受控词表编制新标准的陆续出台和颁布,全国信息与文献标准化技术委员会、《中

图法》编委会等单位都在积极跟踪探索知识组织相关标准的研究,为正在酝酿着启动的《汉语叙词表编制规则》、《多语种叙词表编制规则》这两个标准作修订准备。刘华、沈玉兰、曾建勋的论文《中国、美国和英国叙词表编制国家标准比较研究》从标准组成、标准化对象、叙词表的揭示和组织对象和内容要点等几个方面对中国、美国和英国的叙词表标准进行了比较研究,对他们在标准内容上的共性和差异性进行了详细的对比分析,寻求可供借鉴的经验,并结合我国叙词表实践的发展,对我国叙词表标准的修订提出建议。曹玉强的论文《汉语叙词表主题标引的瓶颈与发展方向》则从主题标引的角度,指出为了解决《中分表》主题标引的瓶颈与问题,网络版的《中分表》是一个最佳发展方向。

伴随着《中图法》第五版修订的进行,国内文献分类领域对当前国内外分类实践进行了广泛研究,内容包括国外分类法修订管理方法、网络版编制进展以及分类法在当前环境下的编制规范、修订技术和电子环境下的应用等。宋剑祥的《网络环境下〈中图法〉第5版修订发展趋向的探讨》着重从6个方面探讨了网络环境下《中图法》第5版修订的发展趋向。常春、吴雯娜的论文《网络时代专业叙词表选词规则实践与讨论》、王世清等的论文《叙词表编制中等同关系获取方法》从叙词表编制的微观视角探讨了叙词表编制过程中的选词规则、词间关系获取等问题。

《军用主题词表》编管会办公室近两年也组织力量研究军用电子分类表编制规则。《军用电子分类表编制规则》作为国家军用标准项目,规定了军用电子分类表编制的基本原则和要求,是指导全军电子分类表编制的技术规范,不久将颁布实施。

3.2 传统信息组织工具的网络化、本体化改造研究

对传统的知识组织工具进行改造的目标首先是使得改造后的网络知识组织系统是机器可理解的、可交换的。本次会议有多篇课题支撑论文对这个问题进行了探讨,体现了对该主题的关注。

深圳大学曾新红承担的国家社科基金项目“基于本体和知识集成实现中文叙词表的升级、共享和动态完善”设计了 Onto Thesaurus 来表示结构规范的中文叙词表(主题词表),以实现其本体化升级和在语义 Web 环境中的共享应用和网络化共建,并在其基础上实现了较为完备的中文叙词表本体共建共享系统(Onto Thesaurus Co-constructing and Sharing System, OTCSS),包括:从中文叙词表到 OntoThesaurus 的自动转换功能、一致性检测推理功能、网络术语学服务功能(供人使用的 Onto Thesaurus-TS 和供应用程序使用的 Web Service OntoThesaurus-API),以及全面深入的网路共建功能(用户界面+修订专家界面)。在所提交的论文中将 Onto Thesaurus 与 SKOS 进行深入比较分析后,提出中文 KOS 的形式化表示不可能完全依赖国外组织制定的规范 SKOS,而必须自行研究制定出一些符合自身特点而又与国际相关标准兼容的规范的建议。

喻菲的论文《〈中国分类主题词表〉网络化研究:从 MARC 到 SKOS》研究了将《中国分类主题词表》(电子版)MARC 数据转化为 SKOS 数据的网络化方法,提出《中分表》“主题词规范数据格式”各字段与 SKOS 词汇之间的对应关系,分析了《中图法》与 SKOS 模型在结构和应用领域上的差异,并提出《中分表》网络版开发同转化出的 SKOS 格式数据相分离的观点。刘丽斌等的论文《〈中国分类主题词表〉的 SKOS 描述自动转换研究》提出了将《中国分类主题词表》电子版自动转换为 SKOS 描述形式的可行性方案,并进行了具体的实践。《中图法》编委会在

会上演示了网络版的功能,并将发行网络版。

贺德方在《〈汉语主题词表〉的回顾与展望》论文中提出中国科学技术信息研究所正在探索一条全新的、基于网络信息的新的编表方法,于2008年设立的“汉语科技词系统建设与应用工程”项目,选择了水利水电专业进行试验,研究在网络环境下叙词表的编制方式和应用领域,为网络时代《汉表》的修订寻求新的编制方法。同时也提出可否将《汉表》转化为本体,使之在本体格式的基础上发挥作用,通过本体促进知识组织系统的整体进步。

3.3 词库集成、互操作研究

互操作(Interoperability)成为国际知识组织领域的研究热点,同时也是网络环境下知识组织必须解决的问题。互操作主要是为解决异构系统和多语言映射的问题。语言障碍和异构系统是用户利用知识信息中的最大阻碍,因此,互操作成为网络知识组织系统重点讨论的一项主攻技术。一般而言,实现互操作有8种方式:继承/仿建、翻译改编、卫星词表、直接映射、共同映射、中心转换、临时列表和协议连接。

跨领域、多来源主题词表的集成服务研究是近年来我国对传统知识组织工具关注的热点。如中国科技信息研究所针对国家科技热点监测的需要,在调研国内各领域主题词表情况的基础上,结合Web Service技术和知识组织技术,提出了跨领域、多来源主题词表的集成服务框架,并对语义集成及关系构词的要点进行了分析,并开始了概念相似度计算方法研究和概念映射工具原型系统的开发。该课题对传统信息组织工具的改造和集成作了许多创新性的工作。

张宇提交的国家科技支撑计划课题——知识组织系统的集成及服务体系研究与实现项目成果《基于概念关系梳理的多领域科技主题词表集成研究》以“交通运输词表”、“汽车工程词表”和“环境科学词表”作为实验对象,进行“新能源汽车”领域的概念集成实验,并着重就实验中的概念关系梳理与整合问题进行实证研究与探讨。

刘华梅的论文《基于情报检索语言互操作的集成词库构建——以教育集成词库为例》探讨利用计算机技术、自然语言处理技术、数学方法等来解决同义词识别、语词匹配、映射关系发现问题,实现不同分类法、主题法间的互操作,进而构建一个以《中分表》为核心的教育集成词库,在此词库的基础上开展各种服务。

3.4 自动标引、自动分类研究

自动标引和自动分类是一个被关注的老问题,这方面的研究很多,但突破性的成果不多,在网络环境下,又再度成为新的研究热点。

自动标引是解决海量信息的快速组织的关键核心技术,无论是网络搜索系统、数字图书馆还是各种数据库都离不开自动标引技术的应用。经典的自动标引理论和方法是基于词频/逆词频的统计方法以及各种改进方法,而傅亮提交的论文《词频与自动标引之关系研究》从自动标引现状出发,通过大量实验数据统计,分析了词频分布规律,明确了词频分布规律对于关键词的选取没有太大作用,揭示了目前自动标引技术过渡依赖词频统计的误区。并指出自动标引技术的发展今后应该避开各种基于频率统计方案的陷阱,着重在语义学和人工智能学的方向上积极努力。更重要的是,人类不应该也不可能绕过某些基础工程的建设,如各类词表(包括词汇、词间关系、词性、语义、语用等),就企图能迅速实现高性能的自动标引效果。传统信息组织方法存在很多困境、一般的机器自动识别、学习的方法因缺乏控制而效果不佳,这就使得

传统信息组织方法与机器自动学习方法的融合尤为必要。

文本自动分类是信息检索与数据挖掘领域的研究热点与核心技术,近年来得到了广泛的关注和快速的发展。自动分类发展到今天,涌现了大量的模型和算法,其中绝大部分是基于统计分类(Statistical Classification)和机器学习技术(Machine Learning Technique)分类器。另一类研究思路是基于知识库的自动分类(以南京农业大学侯汉清研究团队为代表),通过构建一个“分类号(类目词)—主题词—关键词”三者相对应关联的概念语义网络分类知识库,就可以实现自动分类的目标。基于知识库的自动分类的关键在于分类知识库的构建,但是目前分类知识库的建设多采用人工或半人工的方式建立,由于受到人力、物力的限制,这一类的研究都是更多地局限于某一受限领域中进行,因而其在网络上的大规模应用和发展受到了一定的限制。何琳等提交的论文基于《〈中图法〉的多层自动分类影响因素分析》系统总结了基于《中图法》知识库的多层自动分类项目的研究经验,从训练数据、分类算法、类目体系和评估方法等几个方面分析了影响多层自动分类的各种影响因素,同时探讨目前在多层分类中影响正确率的主要障碍及可能的解决方法。

章成志的论文《主题聚类方法的探索与实践》综合了情报语言学、文本挖掘、机器学习、自然语言处理、信息检索等多方面的理论与技术,将信息组织方法中的主题法与数据挖掘中的聚类方法的结合,提出基于主题聚类的方法,并将主题聚类方法应用于主题数字图书馆构建、学科热点检测以及搜索结果聚类等信息服务中。

张运良等提交的论文《事实知识抽取中的语义自足性问题研究》从知识服务的实际需求出发,引入自然语言处理的概念层次网络理论,提出通过语义自足性来处理知识抽取中的知识要素边界问题以及知识对象和知识要素约束问题。

王爱华的论文《情报检索语言与自动标引网络信息资源初探》研究了如何实现网络信息资源自动标引来自动建立网络信息检索系统。

3.5 信息组织理论、方法和工具的行业应用研究

从方法论意义角度分析,信息组织理论、方法和工具的行业应用开始受到关注。面向特定行业进行网络信息组织,可以实现行业内外信息的充分共享,构建行业企业和用户之间的资讯交流平台,为行业外部的知识管理、知识共享作出贡献。面向行业来组织网络信息,将海量的信息范围缩小到特定的行业范围内,从而使网络信息检索工具在检索特定行业信息时得以发挥最大的有效性。郑敏的论文《面向行业的网络信息组织》从宏观上分析了网络上行业信息组织的现状,指出了其中存在的不足之处。在此基础上,探讨了面向行业的网络信息组织思想和方法。李华、康艳的论文《传统信息组织理论在信息服务中的应用实例——〈号百信息分类法〉编制技术初探》是传统信息组织方法应用于现代信息服务业的实践研究,其从分类法编制的实践中得出的经验与体会对行业分类表的编制具有相当的指导意义。郝嘉树的论文《我国公文主题词表分析研究及其发展探讨》对电子政务领域的公文主题词表进行了研究。

此外,衡中青的《中国地方志物产分类方法研究》研究了明、清、民国广东方志物产分类体系表,刘丹和羌丽的《主题图在特色数据库构建中的应用》构建了特种经济动植物主题图,设计了主题图导航系统的总体架构。

3.6 网络信息组织的新技术新方法研究

语义网信息组织、本体构建方法、搜索引擎中的信息组织方法、标注方法等都是近几年研究和应用的热点。2005年前后语义网、本体的研究热潮之后,随着研究的深入,人们对网络信息组织领域的研究逐渐回归理性。人们认识到语义网的瓶颈在本体,本体构建的瓶颈在于领域知识库的构建,因而回过头来发现传统知识组织工具(分类法、主题词表)是本体构建的基础或是改造源,是一条捷径。因此,兼容改造传统的叙词表、分类表研发知识本体、语义网络、主题图等语义工具成为我国网络知识组织工具研发的总趋势;集成知识库的构建、传统知识组织系统的语义描述是其中的较受重点关注的研究方向。

本次会议提交的论文中,范炜的《分面思想的发展路向分析》在对信息组织领域的分面分类研究进展系统梳理的基础上,归纳出分面分类在当前主要被用于构建支持搜索、浏览和导航的底层架构,从而发挥信息组织架构在人们探索性搜寻活动中的辅助作用。提出分面分类研究与实践主要处于信息组织、信息检索以及信息交互三个领域的观点;并立足于信息构建的视角(内容组织、探索性搜索导航等),提出当前分面分类研究的一些重要议题,为传统信息组织思想的研究和发展提供了思路,体现了信息组织向信息构建应用拓展的理念。本次会议特别设立“Web2.0信息自组织研究”这一分主题,旨在于此。另外,本次会议论文中,韩建新的《Folksonomy 检索初探》、包冬梅的《标注行为、标签及其组织模式的优化——以个人学术空间为例》、倪皓的《大众标注的性能及其在图书馆中的应用》、王玉媛的《大众分类与语义网》都是有关自由分类法的探讨,表明研究已经开始向寻求大众分类的优化与适度控制的方向拓展。罗昊的《博客信息自组织模式研究》表明网上庞大的个人信息的组织应该引起我们的关注。

3.7 数字图书馆知识组织系统研究

无论是传统图书馆,还是数字图书馆都需要对信息资源进行有序的组织。在网络环境下,对知识组织系统的要求包括可扩展性、可伸缩性、可互操作性,以及最为重要的一点——简单易用性。

随着数字图书馆研究的不断深入,如何将知识组织工具、方法贯穿应用到数字图书馆中的研究也不断深入。广义上理解数字图书馆知识组织系统,应包括数字资源的知识组织系统构建、自丰富、互操作、术语服务基于知识组织系统的自动采集、自动分类、自动标引及分类导航、智能检索的方法。如何基于传统的知识组织工具和方法,对其进行网络化改造,使其可以应用于数字图书馆和网络环境下,构造数字图书馆的知识组织系统,实现数字信息资源的自动组织,以支持概念检索和知识服务。

卜书庆的论文《基于〈中分表〉的国家数字图书馆知识组织思考》提出了基于《中分表》的国家数字图书馆知识组织系统的建设及应用构想方案。包括基于《中分表》构建“综合性数字资源的知识组织体系”,基于“综合性数字资源的知识组织体系”的领域知识本体构建,基于数字资源的知识组织体系的“国家数字图书馆学科信息门户”建设,基于《中分表》创建多结构知识组织系统的维护系统。

姚慧君、徐庆有的论文《融于数字图书馆的知识组织系统模型构建探究》介绍了数字图书馆知识组织系统(DL-KOS)的类型与功能,提出了数字图书馆知识组织系统的模型构建方法,包括采用元数据进行数字图书馆的知识组织,集成分类法、主题词表和语义元数据构造数字图

书馆知识组织系统,采用专家系统进行数字图书馆的知识组织。

3.8 《信息组织》课程研究

信息组织课程建设也就是 10 多年的时间,近几年,又得到了进一步的发展,中外高校课程体系的信息组织课程的核心地位更加突出,已成为各院校的核心课程。

美国多数图书馆学信息学研究生院从 20 世纪 90 年代就开设了信息组织类课程。我国各高校近 10 年里都开出了信息组织类课程,从各院校网站上浏览可以得知,如北京大学信息管理系、武汉大学信息管理学院、南京大学信息管理系、中山大学资讯管理系、中国人民大学信息资源管理学院、湘潭大学公共管理学院、郑州大学信息管理系等都开设了信息组织课程。从总体或发展趋势来看,信息组织课程已经走向成熟和比较稳定。

戴维民的论文《信息组织课程建设与学科发展》从中外大学图书情报和信息管理类专业的信息组织课程的建立、发展和教学内容的选择和安排,分析了作为学科专业领域的信息组织的发展现状和路向,同时揭示了信息组织课程建设与学科发展的密切关系。

4 祝贺张琪玉教授八十寿辰

2009 年 6 月 7 日是张琪玉教授八十大寿,南京政治学院上海分院军事信息管理系为先生举办了隆重的庆祝和座谈会,大家共同畅谈张琪玉教授的优秀的学术品格,教书育人的高尚师德,严谨的治学学风,丰富的学术思想,永远不息的开拓精神,以及在情报语言学以及索引学等学术领域的杰出贡献。

张琪玉教授作为情报语言学的开拓者,本次研讨会特设“张琪玉情报语言学思想研究”这一专题,目的是了解和学习张琪玉先生一生热爱图书馆事业、致力于情报语言学的研究和探索、五十年来笔耕不辍的学术追求品质,学习张老先生的学术研究方法。张琪玉教授作了自己三十多年从事情报语言学教学研究的体会报告。

张琪玉教授从 20 世纪 70 年代开始以自己的思想,自己的路径、自己的语言、自己的方法开拓情报语言学领域,高屋建瓴地提出情报检索语言理论研究的新方向、新范围、新方法。集中反映张教授情报语言学学术思想的学术著作《情报检索语言》及其增订版《情报语言学基础》,被学界公认为经典之作,先后入选“20 世纪中国学术名著精华目录”,“20 世纪中国图书馆 100 件大事”,位列“20 世纪中国图书馆学最重要的 10 部文献”的第二位。

张琪玉教授在《我研究情报语言学的若干心得和收获——自述学术思想》会议文章中描述了他几十年从事情报语言学的感受,读来很受启发和教益。

张琪玉教授说:“我有一个基本观点,认为各种情报检索语言的基本原理是一致的,它们只是在表达各种概念及其相互关系时在解决对它们提出的那些共同要求时所采用的方法不同,才形成了不同的类型和语种。因此,对分类检索语言、主题检索语言和其他情报检索语言以及自然语言在情报检索中的应用问题进行统一研究,可探索它们影响检索效率的共同规律和有效的改进途径。提高检索效率是情报语言学研究的目的和核心问题。”

“我的学术研究,以情报语言学(包括索引学)为范围。对于情报语言学,可以说是‘情有独钟’。自 1980 年之后写的几百篇著作,几乎全部是属于情报语言学的。”

“我认为,人生很短暂,一个人不能做完一切想做的事,所以只能去做最必要的、最有意义

的并且能够做到的事情。做学问,要有个专业范围,在专业范围内力求深而广。”

“我体会,研究的成效,在很大程度上取决于研究方法。”

徐晏早从一个老同事的角度简要概述了张琪玉教授的情报语言学思想与成就。曹树金等在《中国当代索引学的精髓——张琪玉教授的索引学思想研究》中总结了张琪玉教授在索引学研究的学术思想特征,包括成功将情报语言学应用于索引学研究;强调索引研制的创新性和计算机化;既着眼于宏观研究,更侧重于微观探讨;积极推广优秀索引作品等。

为了庆祝张琪玉教授的八十寿辰,褒扬张琪玉教授的学术成就和贡献,许多学术团体、高校、业内同行和学生纷纷发来贺信和贺电。在张琪玉教授生日座谈会即专场报告会宣读了部分贺信贺电,听来令人感动,以下摘录一些文句诗篇,以飨同行。

中国图书馆学会:

“张琪玉先生几十年如一日辛勤工作,为我国的图书馆事业和图书情报学专业教育的发展做出了巨大贡献。”

中国索引学会:

“先生不但学术成就硕果累累,而且长期投身教育事业,培养了一大批后起之秀,可谓桃李芬芳,嘉惠图林。”

全军军用主题词表编管会办公室:

“您是我国图书情报界的老前辈,是情报检索语言学理论的开创者。几十年来,您著书立说孜孜不倦硕果累累,教书育人诲人不倦桃李满天下,广受业内同行的赞誉和尊重。”

上海图书馆上海市图书馆学会:

“先生淡泊名利的高尚品格,宽厚待人的处世之道,诲人不倦的育人心,严谨科学的风范,孜孜以求的学术精神,为图情界之师表。”

《中国图书馆分类法》编委会:

“张琪玉先生在长期的图书馆实践和教学中,潜心情报检索语言理论研究、探讨情报检索语言的发展,其代表作《情报检索语言》(后改名为《情报语言学基础》)为我国情报语言学的开拓之作;其主持或参与编修的《中国图书馆图书分类法》、《中国分类主题词表》、《汉语主题词表》等工具书在我国图书馆信息领域得到了广泛的应用;其近作《图书内容索引编制法——写作和编辑参考手册》为我国索引工作者提供了编制图书索引的实用指南,深受好评。”

北京大学信息管理系王余光主任:

“张先生1954年毕业于我系,长期在图书情报学界从事教育工作,培养了一大批图书情报界精英。学术上,主张对分类法、主题法和其他情报检索语言以及自然语言在情报检索中的应用问题进行统一研究,以探索它们影响情报检索效率的共同规律和有效的改进途径,在情报语言学的理论建设上有卓越成就。”

武汉大学信息管理学院陈传夫院长:

“为建设和发展军队图书情报教育事业,张老毅然到贵系工作,率领贵系迅速跻身于我国图书情报教育的前列。也有力地推动了我国图书馆学、情报学教育与科研事业的发展。”

南京大学信息管理系:

“张琪玉先生是我国著名的图书馆学家,为我国信息检索语言学的研究做出了重要贡献,是该学科的开拓者。他对信息检索语言以及自然语言在信息检索中的应用问题进行了大量研究,他所作的开创性工作对以后的研究发挥了极为重要的指导意义。”

中山大学图书馆馆长广东图书馆学会理事长程焕文教授：

“张琪玉教授是我国杰出的图书馆学家和情报学家。”

“半个世纪来，张琪玉老师在分类法、主题法等学术领域刻苦钻研，以一丝不苟、严谨科学、勇于开拓的治学精神，开创了崭新的情报语言学学科，成为我国情报语言学的卓越奠基人，为图书馆学和情报学的发展做出了彪炳史册的贡献。”

张琪玉教授指导的研究生中山大学资讯管理系曹树金主任：

“您是我国情报语言学的宗师和当代索引学的泰斗，是我国最杰出的图书情报学家之一。您开创的情报语言学不仅在我国图书馆学和情报学中独树一帜，而且适时地引领了我国图书馆学和情报学的转型和跨越，为图书馆学和情报学的发展作出了无与伦比的贡献。

您亲自培养了数以千计的情报语言学人才，数以万计的学人受到您所创立的情报语言学的熏陶。您呕心沥血，为图书情报事业培养了一批又一批国家栋梁，桃李满天下，享誉海内外。

几十年来，您兢兢业业，将全部身心倾注在图书情报学的研究和教育上，您严谨治学，一丝不苟、勇于开拓，令人景仰；您为人师表、诲人不倦、奖掖后学，垂范学界；您一心追求真理，淡泊名利得失，实为学人楷模。”

上海图书馆馆长上海科技情报所所长吴建中先生在张琪玉教授八十大寿之前赋诗一首，表达对张教授的祝福。

信息引领胜利戟

——为张琪玉教授八十大寿而作

三军独树一面旗，
信息引领胜利戟。
人才强学是基石，
承前启后人心齐。

张琪玉教授的学生，现任中国证监会信息中心副主任李为女士2009年6月11日在飞赴上海会议的航班上在手机上写下了天上诗篇。

恩师张琪玉先生八十大寿祝辞

检索语言，我师唱响。
武大初识，多年既往。
桃李芳香，学界敬仰。
网络时代，更有发展。
值此大寿，感恩培养。
学术会议，新人成长。
与时俱进，定创辉煌。

张琪玉教授的学生李国秋现在丹麦做学术访问，在张教授八十大寿之际发来创作的诗作一篇，以表弟子的敬仰之心。

祝张老师八十大寿

二十年前师张老,德才风采脑屏中。
三地勤耕京沪汉,桃李纷菲满苍穹。
鸿篇巨著创百代,育人美誉震长空。
龄高耄耋知何似?屹立图岭一巨松。

5 关于未来研究思路与路向的思考

在网络环境下,信息组织的未来趋向始终是研究和关注的热点问题,此次会议有多篇文章论及了此议题。薛春香、侯汉清的论文《知识组织系统及其描述与评价初探》提出知识组织系统是网络环境下发展起来的更高层次的知识组织语义工具,并且向智能化、精深化、专业化的方向发展。陈俊凤、何晓萍的论文《情报检索语言在网络环境中的发展趋势探究》提出受控语言与自然语言相结合用于组织和检索网络信息是未来发展的必然趋势。

5.1 变革研究思路

很具巧合意义的是大会报告推荐了两份关于情报检索语言的六点建议。一是南京农业大学侯汉清先生推荐的 Keven 在博文中提出的情报检索语言应脱胎换骨六点建议^①:(1)编撰、修订的机制和方式已经不合需要了;(2)技术背景和应用环境变了,应用目的肯定有所不同,特别是对于 Web 环境下的信息系统,想要起到规范控制的作用,设计思路一定得跟着变;(3)与第一点相关的,开放开放还是开放问题;(4)我们的分类主题词表要适应网络时代的要求,在设计原则和技术方面可能也可以进行许多改进;(5)包括分类法、主题法、知识本体等各类具有一定规范的知识组织体系是数字图书馆建设的重要依据和标准;(6)现在不光要有电子版,更重要的是应该有网络版。

另一个是中国证监会信息中心李为女士推荐的该中心的计算机博士初壮关于对情报检索语言发展方向的六点建议:(1)词表一定要能够动态地更新,能及时反映所应用领域的最新需求;(2)词表一定要用计算机语言去实现,一定要考虑如何被计算机信息系统所用;(3)传统的词表和目录体系,实际上是一个特定领域信息的“全局遍历”,但是在计算机时代,全局遍历的需求被弱化,而通过检索和对检索结果的“遍历”反而成为重点;(4)既然词表的使用对象是计算机系统,词表的体系结构,将不再是单线索的而是多维的,从单一维度的目录体系发展到多维向量体系;(5)检索语言,要实现从面向专业人员,转变到面向大众;(6)检索技术的应用,还要考虑实现过程的效率和成本。

前六点建议出自业内同行,是一篇博文,用网络语言写成,后六点是计算机专业人士的思考,给上级的建议,用语平和。但二者的建议都是开门见山,观点鲜明,无论你听起来是否顺耳,这些建议都应该引起我们的重视。我认为他们提出的核心问题就是情报检索语言的现代性和大众性两个问题,这需要我们改变思路和方法。网络时代为情报检索语言和当今的信息

^① 情报检索语言脱胎换骨的六点议论: <http://www.dlresearch.cn/keven/index.php/archives/488> (原博文发表于2005年11月24日)