

戴维民

## 中国情报检索语言 50 年研究论纲

**摘要** 50年来,我国情报语言学逐步走向完善和成熟:发表和出版了上万篇(种)专业文献,形成了一支优秀的核心研究者队伍,建立了情报语言学学科理论体系和专业课程教学体系。在情报语言学文献、作者、分类法、主题法、一体化词表、标准化与兼容化、新技术应用等的实践与理论研究方面都取得了可喜成绩,并在面向 21 世纪的新型情报检索语言的研究中取得了前瞻性的成果。

**关键词** 情报检索语言 1949~1999 年 研究 综述

**分类号** G354.2

**ABSTRACT** In the last fifty years, studies in information retrieval languages are developing towards perfection and maturity, which are reflected in the following aspects: about ten thousand titles of research papers/ works have been published, a group of researchers has been formed, a theoretical system and educational curricula have been established. Some good results in the research of new information retrieval languages oriented to the twenty-first century have been obtained.

**KEY WORDS** Information retrieval language. 1949-99. Studies. Review.

**CLASS NUMBER** G354.2

### 1 引言

50年来,我国情报检索语言研究日益繁荣,情报语言学逐步走向完善和成熟。其主要标志是:发表和出版了上万篇(种)的专业文献,形成了情报语言学基础著作和核心文献;形成了一支优秀的核心研究者队伍,涌现出了一大批造诣很深的学科开拓者和学术带头人;建立了情报语言学学科理论体系和课程教学体系;编制了适应各类文献机构、各种文献类型的综合性或专业性的分类法、主题词法和分类法主题法一体化词表;制定和颁发了系列的检索语言编制标准、标引标准和作为标准的分类法和主题词表;在引进计算机技术,研究标引和检索技术等方面取得许多突破性进展,自然语言检索技术研究展现了良好的发展前景;对面向 21 世纪的新型情报检索语言和情报语言学开展了积极研究,取得了一些具有前瞻性的成果。

### 2 情报语言学文献研究

情报语言学作为一个比较成熟的分支学科,其中一个显著的标志就是发表和出版了一批数量可观的研究文献,其中不乏经典之作,形成了本学科的基础文献和核心文献。

根据张琪玉编制的《情报语言学文献库》(电子版)按年度统计,1949~1999年共计9343篇(种),如果收全1998、1999年文献,50年中总文献量可望达到万篇以上。在50~70年代的30年中,情报语言学文献数量是很有有限的,50年代为164篇(种),60年代为220篇(种),70年代为158篇(种),说明建国后的30年中,情报检索语言的研究是仅仅处于一个起步阶段,研究的问题面也很窄。特别是在建国之初和文革动乱的一些年度中出现了空白或文献量极少的情况,说明建国之初百废待兴,10年动乱,学术研究几乎处于停滞状态。

情报语言学作为一个分支学科,其发展和繁荣开始于80年代。80年代前五六年中,文

献数量成百地递增,至 1985 年,年文献量已达到 507 篇(种),整个 80 年代的文献量达到了 4360 篇(种)。

90 年代在 80 年代的基础上继续保持高度繁荣,在一个很高的平台上保持快速发展,在 1998、1999 年文献收录非常不全的统计下,总数也达到了 4397 篇,可见情报语言学在 90 年代已走向成熟,而且保持着强劲的发展势头。

仅从文献数量看,我国情报语言学学科真正的发展历史不过 20 年。从 80 年代开始,情报检索语言从经验型和实用性的圈子中走出来,开始其学科理论的构造,搭建其学科框架,使情报语言学作为一个分支学科逐步走向完善和成熟。

我国情报语言学论文的主要情报源是(括号中数字是发文量):《图书馆学研究》(344)、《中国图书馆学报》(234)、《情报理论与实践》(230)、《图书馆建设》(219)、《图书情报工作》(218)、《科技情报工作》(193)、《现代图书情报技术》(183)、《图书情报知识》(181)、《江苏图书馆学报》(180)、《图书馆杂志》(178)、《大学图书馆学报》(173)、《图书与情报》(159)、《四川图书馆学报》(143)、《图书分类论坛》(137)、《图书馆理论与实践》(137)、《山东图书馆季刊》(126)、《情报资料工作》(122)、《情报科学技术》(118)、《情报学报》(111)、《云南图书馆》(108)、《情报业务研究》(107)、《图书馆学刊》(103)、《图书情报论坛》(102)。以上登载 100 篇论文以上的 23 种专业期刊,可以确定为情报语言学的核心期刊。

孙勇和王秀玲曾对 1949~1995 年我国情报语言学著作进行了统计分析,共 143 种,其中编著 104 种(占 72.7%),译著 39 种(占 27.3%)。1995 年以后,也陆续有著作出版,因此至今,著作总数应在 150 种以上。

根据情报语言学著作自身的影响,以及对情报语言学学科建设的影响,以下 10 部著作可认为是情报语言学的核心著作:《情报语言

学基础》(张琪玉编著)、《现代西方主要图书分类法评述》(刘国钧著)、《图书分类学》(白国应编著)、《图书分类》(李严、史永元、张函、侯汉清、马张华编著)、《图书分类学》(周继良主编,周绍萍、俞君立、张燕飞编著)、《图书馆分类工作手册》(侯汉清,王荣授主编,陈树年、戴维民副主编)、《汉语主题词表标引手册》(钱起霖主编)、《主题法的理论与标引》(刘湘生编著)、《情报检索与主题词表》(丘峰著)、《主题法导论》(侯汉清、马张华主编)。

在情报语言学著作中,还出版了一批文集,主要有:《刘国钧图书馆学论文选集》、《杜定友图书馆出版学论文选集》、《张琪玉情报语言学文集》、《当代分类法主题法索引法研究》(侯汉清)、《情报检索语言综论》(戴维民)、《联机环境中的情报检索语言》(曾蕾)、《阮冈纳赞理论及其研究》(叶干军)等。综合性文集主要有:《情报检索语言论文选》、《图书分类教学参考资料》、《图书分类学文集》、《文献分类学论文荟萃》等。

国外情报语言学系统著作在我国翻译的约在 40 种以上,其中主要的有:《情报检索理论概述》([苏]切尔内著)、《情报检索词汇控制》([美]兰开斯特著)、《图书分类理论》([英]布坎南著)、《冒号分类法理论与实践》([印]塞克德瓦著)、《冒号分类法解说及类表》、《需要以分面分类作为一切情报检索方法的基础》(英国伦敦分类法小组著)、《分面分类法——专业分类表的编制和使用指南》([英]维克里著)、《主题检索系统设计选择》([美]米尔斯蒂著)、《国外情报检索语言研究——分类主题一体化的理论与实践》(吕其苏等编译)、《情报组织与利用》([美]索尔格尔著)、《文献标引手册及实例》、《图书资料的分类》([日]宫坂逸郎、石山洋等编)、《图书分类法史略》(第 2 卷)([苏]沙姆林著)。

### 3 情报语言学文献著者研究

在 50 年的发展中,我国情报语言学形成

了一支数量庞大、有研究核心的研究者队伍,这是学科发展和建设的重要因素。

根据《情报语言学文献库》的作者的统计分析,50 年来情报语言学的个人作者总数为 4516 人,其中发文 1 篇的作者数为 3153 人,发文 2 篇以上的 1363 人,发文 10 篇以上的 98 人,发文 20 篇以上的 28 人,其中:张琪玉(146)、侯汉清(136)、白国应(88)、戴维民(85)、皮高品(62)、刘国钧(51)、杜定友(50)、钱亚新(48)、陈光祚(44)、刘湘生(38)、叶千军(38)、傅兰生(32)、蔡振华(32)、史永元(32)、丘峰(31)、陈晓华(30)、李严(30)、李锡初(28)、刘延章(26)、高崇谦(25)、曹树金(24)、吴声亮(24)、俞君立(24)、曾蕾(24)、王永成(24)、赖茂生(23)、陈树年(20)、黄恩祝(20),这 28 人可认为是情报语言学的第一核心作者。

情报语言学的核心作者群是情报语言学学科的主要建设者。其中有些堪称情报语言学的奠基者和开拓者,是情报语言学学术带头人。例如:刘国钧、杜定友、皮高品、钱亚新和张琪玉等。

#### 4 分类法理论与实践研究

建国之初的《东北法》和《山东法》由于编写及时,从而在我国当代分类法史上占有重要地位。50 年代掀起了我国分类法编制的— 高潮,共达 30 部之多,基本满足了图书分类工作的需要,这期间比较有影响的分类法有:《广东人民图书馆图书分类法》、《中国人民大学图书馆图书分类法》、《中小型图书馆图书分类简表》、《浙江图书馆图书分类新表》、《中国科学院图书馆图书分类法》、《武汉大学图书分类法》等。此外,还对有些旧分类法进行修订后继续使用,如金云铭 1928 年编制的《中国图书分类法》于 1955 年修订;刘国钧 1929 年编制的《中国图书分类法》由北京图书馆于 1957 年修订,并在北京图书馆使用。

从 60 年代开始,图书馆界极力寻求解决

统一分类法问题,1960 年编制《大型图书馆图书分类法草案》就是一次重要努力,虽然编制工作半途而废,但为 70 年代《中国图书馆图书分类法》的编制奠定了基础。1975 年《中图法》第 1 版的出版,开始了我国统一分类法的新时代。80 年代以后,我国分类法实际上已基本走向统一,从分类法的编制转到及时修订分类法上来。但其间编制的《标准文献分类法》和《中国档案分类法》等是针对不同文献类型的重要分类法。

50 年中,在理论和实践应用方面具有重要影响的分类法是:《中国人民大学图书馆图书分类法》、《中小型图书馆图书分类表草案》、《中国科学院图书馆图书分类法》、《大型图书馆图书分类法》、《中国档案分类法》、《中国图书馆分类法》。

近现代分类法起源于西方国家,因此引进和借鉴国外的先进理论和技术对我国分类法的理论研究和新型分类法的编制具有直接意义。

建国之初,我国新型分类法的编制思想主要是借鉴前苏联的做法。翻译了一些苏联的分类著作和几部分类法。安巴祖勉、托罗帕夫斯基、杰斯林科都成为了我国比较熟知的专家学者。《苏联图书十进分类法》(改订本)、《十进分类法》等被翻译成中文供参考使用。

由于编制新型分类法的需要,开展了大量的理论研究,对分类法的编制技术也进行了有益的探讨,可以说涉及到分类法理论和技术的各个方面。

1950 年 6 月 2 日,文化部副部长郑振铎亲自主持召开了图书分类法座谈会,讨论分类法编制的理论与技术问题,这是我国建国后召开的有关分类法的第 1 次学术会议。此后召开的各种学术会议和工作会议比较多,其中有较大影响的是“铜陵会议”和“天津会议”。

1985 年在安徽铜陵召开的“《中图法》(第 3 版)修订研讨会”就《中图法》和《资料法》中如何扩大使用组配方法和是否要为综合性

科学立类问题进行了研究。1991 年 1 月在天津召开的“全国《资料法》分面改造学术研讨会”主要是为《资料法》分面改造决策提供方案及参考意见。会议对《资料法》分面改造的必要性及重要性、分面改造的路向和原则,以及技术方法等 3 方面的问题进行了研讨。

我国分类法理论和编制技术研究的重点主要集中在如下方面:(1)图书分类法体系的建立,包括基本部类及其序列的确立,基本大类的划分,马克思列宁主义、毛泽东思想的类目的设立,社会科学和自然科学部类的序列,按国家划分类目,依人列类,区分观点,传记类图书的集中与分散,综合性科学列类等方面的问题;(2)体系分类法的研究,包括:“集中与分散”的矛盾以及处理这个矛盾的方法,分面分析理论及体系分类法的分面改造技术,复分表交替类目、注释、参见等的研究;(3)关于分类法思想性、实用性、科学性“三性”原则的研究,以及分类法提高情报检索效率的基本方法研究;(4)分类标记的研究,包括标记符号、标记制度,以及双位数法、八分法、借号法、字母标记法等方法技术;(5)分类法标准化问题;(6)图书分类自动化及机读分类表问题;(7)关于分类目录的研究,包括分类排架和组织分类目录职能的协调、分类目录与主题目录的关系及分类目录前途问题;(8)关于图书分类实践若干问题研究,包括各类、各科文献的分类,新学科文献分类,同类书区分,图书改编等;(9)关于基本理论研究,包括分类法的概念逻辑基础和知识分类基础,图书分类历史,图书分类学(文献分类学)等;(10)关于国外图书分类法及其理论和技术的研究等。

## 5 主题法理论与实践研究

1964 年 5 月,我国航空部编印《航空科技资料主题表》,它是我国建国后编制的第 1 部主题词表,但属标题表性质。至 1971 年 12 月,该词表第 2 版出版时,才标志着我国叙词语言有了进一步的发展。

《汉语主题词表》的编制出版是我国叙词语言发展的重要里程碑。“汉表”是作为“七四八”工程,即“汉字信息处理系统”的配套项目编制的,它的编制锻炼了编表队伍,为我国叙词语言的发展作了人才准备。70 年代编成的叙词表还有《电子技术汉语主题表》、《原子能科技资料主题词典》、《机械工程主题词表》等约六七部。

80 年代是我国叙词语言的大发展时期,编制的叙词表至少在 60~70 部以上。因此,80 年代中期,曾有一场分类目录与主题目录之争,有人认为主题目录要取代分类目录。由于叙词语言的发展,全国汉语叙词表发展方向研讨会于 1988 年 12 月在北京召开,以期研讨叙词语言有关技术问题。从 80 年代后期开始,寻求编表模式和编表技术的突破,其中分类法主题法一体化和运用计算机技术编表是重点,均取得了实用性成果。

90 年代以来,我国叙词表编制的大发展势头不减,编制的叙词也有几十部之多,其中最具代表性的有《军用主题词表》及其系列专业词表、《中国档案主题词表》等。1996 年 7 月在北京召开第 2 次全国分类法、叙词法发展方向研讨会,不仅展示了 90 年代的理论成果,也展示了实践成果。

到目前为止,我国已编制的叙词表在百部以上,几乎覆盖了各专业和各种文献类型,构成了我国叙词语言体系。其中最具代表性的大型词表是《汉语主题词表》和《军用主题词表》。

主题法理论与技术研究主要有如下方面:(1)叙词表的基本原理的研究。确立概念组配的基本原理,并且综合了体系分类法、组配分类法、标题法、单元词法以及关键法等多种检索语言的原理和方法。(2)叙词表的宏观结构和微观结构的优化。在宏观结构上打破了“汉表”所确立的“五表齐全”式的宏观结构,可根据词表应用的需要选择结构,特别是分类法主题法一体化的对照表、分面叙词表的采用,使

汉语叙词表的结构有了新的突破。《国防科学技术叙词表》和《军用主题词表》的等级关系全显示的应用,极大地优化了词表的微观结构。(3)计算机辅助技术的应用,提高了编表效率,保证了编表质量,而且使词表的编、管、用一体化。在 90 年代编制的词表都运用了计算机编表技术。(4)词表编制和使用的标准化。制定了一系列标准,使词表纳入了规范化的范畴,词表的标准化也带来了词表的兼容化的进展。(5)叙词表与自然语言的接口及结合使用,以及运用叙词表控制手段编制自然语言检索的后控制词表。

## 6 一体化词表理论与实践研究

分类法主题法一体化是当今情报检索语言主要发展趋势之一。从 60 年代初探索开始,到 90 年代,我国共发表了 100 多篇研究和评论文章,编制了近 20 部一体化词表。一体化词表成为我国情报检索语言的一个重要组成部分,编制技术和理论也走向成熟。在《主题法导论》和《情报语言学基础》等教材中,一体化词表都作为一章写入教材体系中。

早在 60 年代,刘国钧、杜定友就对分类法和标题法进行了比较,并提出了两者相结合的思想。

1976 年丘峰发表《主题法与分类法》一文将分类法和主题法各种类型进行了比较。1983 年侯汉清发表的《也论分类法和主题法的异同》则注重同时从相同和不同两个方面对分类法和主题法进行比较。对分类法和主题法更系统、更深入的比较研究是张琪玉发表的一系列论文和著作。在其《情报检索语言》一书中在论述到每种语言时都设有专节进行比较研究。

1981 年侯汉清在北京大学图书馆学系以题为《论分类法和主题法的一体化》完成了他的硕士学位论文。此后,作者又先后撰文评介了国外几部具有代表性的分类主题一体化词表。

一体化词表有多种编制模式,如改造分类

表、改造叙词表、编制对照索引、编制分面叙词表等。我国一体化词表编制模式主要是后两种。

北京师范大学图书馆学系杨丹于 1984 年、武汉大学图书情报学院的邓顺国于 1986 年进行的研究为编制《中国分类主题词表》提供了理论和实践依据。

1987 年由北京图书馆图书馆学研究部和武汉大学图书情报研究所共同提出编制《中国分类主题词表》的计划。经过 5 年的编制于 1993 年完成。

分类主题一体化最理想的模式是分面叙词表。1983 年太原机械学院熊兴成编的《常规武器工业分面叙词表》,是我国编制分面叙词表的首次尝试。

1986 年底在北京召开的全国社会科学情报工作会议决定,由中国社会科学院文献情报中心负责筹备组织全国力量编制一部社会科学情报检索语言。1987 年 3 月约请全国部分专家对这一方案进行论证,倾向于编成分面叙词表。会后由侯汉清和戴维民编制的《教育分面叙词表》成为编制分面叙词表的示范表。虽然《社会科学检索词表》历时 10 多年才编成出版,但期间对分面叙词表的编制技术却进行了深入研究,同时有多部专业性的分面叙词表编成出版并投入应用。

## 7 标准化理论与实践研究

1979 年全国情报文献工作标准化技术委员会成立,其中第五分会是主题标引分技术委员会。此后 20 年来,情报检索语言标准化及标引标准化取得了一些显著进展。

50 年代的《人大法》和《科图法》由于其编制技术的先进,使用的面比较广泛,影响比较大,但还达不到统一分类法的目标。1975 年《中图法》的问世开创了分类法的新时代,它被各类图书馆广泛使用,从此不再讨论编制新的综合性分类法问题。1980 年《中图法》第 2 版修订出版以后,全国各级各类图书馆有 96 %

使用它。事实上它已成为一部标准分类法。“汉表”出版以后,其所确立的词表结构体系和词汇基础都得到广泛的认可,为其他词表的编制起到了规范作用,具有标准化意义。

从 80 年代开始,标准化研究和标准的制定进程加快,提出了系列标准草案和送审稿。根据这一背景,文标会五分会等于 1989 年 10 月底在成都召开了全国分类法与主题法标准化研讨会。会议的目的是推动我国图书情报和档案系统检索语言体系逐步走向标准化。会议的主要内容是讨论有关分类法、主题法的 4 项国家标准(送审稿),即《文献分类颜色标识规则》、《中图法》作为国家标准送审稿、《文献分类规则》国家标准(送审稿)、《文献叙词标引规则》国家标准(修订送审稿)。

到目前为止,情报检索语言及标引方面的标准尚有 2 项未批的标准报批稿:《文献分类标引规则》(报批稿)和《文献分类颜色标识规则》(报批稿)。已作为国家标准批准颁布的有 5 项:GB/T13745 《学科分类与代码》、GB3860 《文献叙词标引规则》、GB13190 《汉语叙词表编制规则》、GB15417 《文献多语种叙词表编制规则》、GB15418 《档案文献标引规则》。

90 年代始,各行业和军队也相应制订了一些标准,其中最具影响的是关于军用主题词表编制规则和军事文献标引规则的国家军用标准,具体包括:GB1776—93 《军用主题词表编制规则》、GB2418.1 《军事文献主题词标引规则》、GB2418.2 《军队档案主题词标引规则》、GB2418.3 《军队情报主题词标引规则》、GB2418.4 《军事图书主题词标引规则》、GB2418.5 《军队机关公文主题词标引规则》。

在标准化方面,长期研究和呼吁的是关于书次号的标准化,许多学者进行了积极倡导。实现书次号编制标准化最有效的途径就是编制统一使用的著者号码表。但我国建国后,人们各自为政,编制了几十种著者号码表。其中

《通用汉语著者号码表》是由文标会五分会组织编制的,主要是为了促进我国中文图书同类书排列书次号的规范化、标准化,是一部编制的非常实用的著者号码表。但各图书馆在书次号编制上都有自己的习惯做法,因此,《通用汉语著者号码表》推广是有一定的难度。

外文图书采用的著者号码表基本趋向统一:英文图书采用《克特著者号码表》,俄文图书采用《俄文图书著者号码表》,日文图书采用《植村著者号码表》等。

在我国标准化进程中,也做了大量引进和介绍国际标准和其他国家标准工作,以对我国标准制订起指导和参考作用。

UDC 作为全世界广泛使用的分类法,实际上是一部国际标准分类法。我国 GB7713—1987 规定,科学技术报告、学位论文、学术论文的封面上应尽可能注明 UDC 号。ISO 规定,凡该组织及其成员国颁布的标准文献上必须印 UDC 类号,我国也已执行该决议。《国际专利分类法》是专利领域的一部最通用的分类法,是实际上的标准分类法。英国标准学会为 ISO 编制的《基础叙词表》,企图以标准形式确立其标准性质,我国也翻译了这部一体化词表,名称为《标准文献主题词表》,专门用于标准文献的标引。我国引入的国际标准主要有:ISO2788—1986 文献工作——单语种叙词表编制与修订准则、ISO5693—1985 文献工作——文献审读、主题分析与选定标引词的方法、ISO5694—1985 文献工作——多语种叙词表的编制与修订准则。

与标准化同步,我国情报检索语言兼容化在实践上也取得了丰硕成果。主要表现在:(1)通过制定标准,来规范词表的结构模式和选词标准,如叙词表编制的国家标准和国家军用标准。(2)《中图法》和《汉语主题词表》等综合性检索语言由于其权威性和影响力,同专业性的分类表和词表在实践上实现兼容。(3)通过编制对应表式一体化词表和分面叙词表式一体化词表,实现两种语言的兼容。(4)在一

个系统内,通过有效的组织和提供统一的计算机辅助编表系统,实现系统内部兼容,如《军用主题词表》与 20 部专业词表实现了有效兼容,《国防科学技术叙词表》在编制机读版时,将 11 部词表集成起来,实现兼容。

## 8 情报检索语言新技术应用研究

计算机技术的应用是情报检索语言及标引与检索技术取得重大技术进步的基础。

我国 70 年代开始探讨叙词表的计算机编制和管理技术,但直到 80 年代才有应用成果。中国科学院计算技术研究所从 1983 年 10 月开始研制词表管理系统(TMS 系统),用于编制和管理《计算机科学技术英语词表》。1987 年在南京和上海成功地用计算机软件在 IBM/PC 微机上编制了两部汉语叙词表——《教育分面叙词表》和《教育科学叙词表》。后来的《教育主题词表》、《中国分类主题词表》等都是运用计算机技术编制成功的。有些在编表的同时建立了比较完善的词表管理系统,其中最具代表性的有《军用主题词表》、《农业科学叙词表》和《国防科学技术叙词表》管理系统等。

由文化部立项,山西省图书馆与北京图书馆、中国科学院文献情报中心、中国人民大学图书馆研制的《计算机文献标引对照系统》于 1997 年完成并通过鉴定。它在《中国分类主题词表》基础上予以扩展,实现了《中图法》、《科图法》和《人大法》号码的自由转换,以及三法的号码与《汉语主题词表》的主题词相对应,具有联机标引功能,也可用于检索选词选号。

我国目前开发应用的词表系统中,《军用主题词表》应用管理系统(EMT)、中国农业叙词库及其管理系统、《电子版国防科学技术叙词表》具有一定的代表性。

我国早在 1963 年就开始介绍国外自动标引研究情况,但是直到 80 年代,汉语自动标引研究才开始起步。1980 年 11 月至 1981 年 4 月,陈培久利用 TK—70 计算机和 T4100 汉字

系统建立了“汉语科技文献自动标引试验系统”。80~90 年代,国内学者提出许多自动标引方案。这些方案从类型上划分有形式匹配法、切分标记法、统计法、语法语义分析法、神经网络法等。

自动分类研究开始于 50 年代末,现已达到实用化阶段。自动分类包括自动聚类、自动归类和类号的自动转换 3 部分。1982 年侯汉清与黄刚发表《电子计算机与文献分类》探讨和介绍了国外自动分类的研究情况。到 80 年代中期及其以后,我国有不少学者对自动分类进行了试验和系统开发。如:朱兰娟在王永成指导下进行的试验研究,1995 年金巍在张琪玉和王永成指导下完成的以肿瘤学专业文献为例的中文自动分类系统研究,1995 年苏新宁、徐进鸿、史九林合作研究的档案自动分类方法,1995 年叶新明关于基于《中图法》的中文文献自动分类研究等。我国研制和试验的主要自动分类系统有:辅助归类系统(莫少强,1984)、自动归类系统(朱兰娟等,1986)、辅助归类系统(张炳恒,1989)、自动归类系统(苏新宁等,1995)、自动归类系统(叶新明,1995)、自动归类系统(吴军,1995)、自动归类系统(王永成、张坤,1997)、自动归类系统(刘开瑛等,1997)。

全文检索是自然语言检索的重要方式。我国全文检索系统中以北大方正和东方龙马开发的系统最为出名。世界流行的全文检索模型有布尔逻辑检索模式、概率推理模型和空间向量模型。我国的检索模型单一,主要以布尔逻辑检索为主;各类截词检索是最常见的方式;概率推理方式和空间向量方式,理论研究少,实用的检索系统更少。

后控制词表理论研究日臻成熟,对我国词表的功能、形成等问题的研究带来了新的视角。自然语言检索作为情报检索的一种方式,离不开情报检索的基本原理;后控制词表作为对自然语言的有效控制,为改进联机网络内的检索效果起到了重要作用。专家学者一致认

为,后控制词表是实现受控语言与自然语言结合的有效途径,并对其控制机理、控制程序、检索效率和词表编制方法展开了广泛讨论。

自然语言接口是叙词表中自然语言检索的入口,性能良好的自然语言接口能较好地实现自然语言与受控语言的转换,提高自然语言与叙词语言的兼容度,提供多种查检途径。1998 年通过鉴定的“《军用主题词表》应用管理系统”提供的自然语言接口是通过叙词词素的轮排来实现的,并加入了人工智能的原理和方法,在国内处于领先地位。

## 9 面向 21 世纪情报语言学研究

从 80 年代后期,尤其是 90 年代,许多学者对未来情报检索语言进行了展望和构想。

曾蕾认为在联网环境中情报检索语言有以下发展趋势:情报检索语言词汇的联机显示;多数据库查询中的词汇兼容,采用自由词以提高联机状态下的主题途径查找效率;加强检索语言及检索系统的易用性和透明性。

张琪玉在 1997 年分别在《图书馆杂志》和《北京大学学报》(信息管理系专刊)上发表文章,披露了他关于未来新型情报检索语言的设计思想,即学科—事物概念—组配型检索语言,它是一种集各类情报检索语言以及自然语言之所长的新型检索语言模式。

白国应认为,21 世纪中国检索语言主要有 10 个发展方向,即多样化(从单一化走向多样化)、一体化(分类法主题法一体化)、兼容化、国际化(通用化)、组配化、规范化(标准化)、自动化、自然化(采用自然语言)、专业化和整体化。

张琪玉在 1997 年发表的《世纪之交中国

情报语言学发展之路》一文中,提出了我国情报语言学面临的 8 个方面的研究课题。次年又在《情报语言学领域亟待研究且潜藏较富的课题》一文中,更加明确地提出了我国情报语言学未来需要研究的 5 个方面的研究课题:(1)情报检索语言研究课题,包括体系分类法的改造,叙词表的分面化改造及与自然语言相结合,自然语言接口及对应转换词典;(2)自然语言检索研究课题,包括自动抽词及抽词词典,单汉字检索系统的改进,全文检索系统的改进,后控制词表编制的自动化;(3)网上资源检索研究课题,包括适用于组织和检索网上资源的检索语言的探索,中外主要情报检索语言之间对应转换,中外主要自然语言之间的对应转换;(4)情报语言学基础研究课题,包括知识分类的微观立体网络化,情报语言学与术语学的结合,未来情报检索语言的探索,用户研究;(5)文献标引和检索某些具体问题的研究课题。

总之我们认为,在 21 世纪,情报检索语言仍是发展主流,并将广泛应用。同时自然语言应用范围将越来越广,并在应用中充分吸取情报检索语言的原理和方法,走相互结合、相互融合的发展之路。在情报检索实践中,情报检索语言将不断优化、改进、探索出新型的理想的情报检索语言模式。

戴维民 南京政治学院上海分院(原空军政治学院)信息管理系主任、教授,《文献工作研究》杂志主编,主要研究情报语言学。通讯地址:上海市江湾五角场。邮编 200433。

(来稿时间:1999-11-01。编发者:翟凤岐)