

吕其苏 等编译

国外情报检索 语言研究

——分类主题一体化的理论与实践



社会科学文献出版社

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、侯汉清、朱铁声等编译



社会科学文献出版社

1989年 北京

2553/16

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

社会科学文献出版社出版、发行

(北京建国门内大街5号)

新华书店经销 牛栏山印刷厂印刷

787×10921/32开本 9.25印张 206千字

1989年7月第一版 1989年7月第一次印刷

印数0001—2600

ISBN7-80050-089-6/H·1 定价：2.95 元

前 言

目前世界上情报检索语言正进行重大变革。这是由于现代科学技术的高速发展,各种文献的迅猛增加,新学科和边缘学科的不断出现,新概念和新命题不但层出不穷,而且处于不断发展变化之中,传统的体系分类法的树形结构模式已不能适应形势的需要。因为十九世纪确定下来的近代分类法的基本依据是传统的知识门类的稳定性,学科体系的等级性和概念从整体到局部的排列原则。但是现代分类法,在原有的基本大类中,出现了与传统学科相并列的跨学科的主题类目;由于大量出现了属种之外的其他关系,建立在属种关系上的等级性特征不再占主要地位;概念排列的原则又恰恰是从局部开始,特别是分类的对象已不仅是传统的知识门类和学科,而是所有的“语义单位”。从而使传统分类法赖以存在的基本原则都不同程度地受到了冲击。同样,传统的字顺主题法也发生了深刻的变化,现代主题法(叙词法)越来越多地吸取分类因素。这样一来,分类语言大量增加了主题因素,主题语言也大量吸收了分类因素,结果使这两种情报检索语言相互结合,向一体化方向发展。

一体化的核心是分类语言和主题语言的兼容互换问题。然而它们能否做到完全的兼容互换呢?从认识论上看,是可以做到的。传统体系分类法是根据学科的某种基本属性,按其体系组成单维的等级系统。它只能反映知识结构的单维关

系。分面分类法是根据学科的主题内容，按其多种属性（若干个范畴面）组成多维的组配系统。它反映知识结构的多维关系。可见分面分类法与传统分类法在其功能上大致可以体现出事物的一般和个别的关系。两者所反映的是同一个客观实体，只是反映的专指性和完备性有所不同。多维关系必然包容单维关系。这就是分类主题一体化认识论的基础。

经过多年来的理论和实践证明，分类主题一体化最理想的模式是分面叙词表。它的突出特点是能把分类语言和叙词语言的概念系统、参照系统、标识系统以及索引系统实现相互兼容，能将两种以上的情报语言融为一体，并能充分发挥它们各自的长处，以实现最佳的整体效果。叙词表的每一个叙词或几个叙词的组合，即为分面分类表的一个范畴面的类目。反之，这些类目又是叙词表的一个或几个叙词的组合。通过这种双向联系，使分类标引和主题标引可一次完成，从而提高了标引工作的效率。同时，用户也可以经济、有效地同时进行分类查询和主题查询。

分面叙词表出现于70年代初，是近二十年的事情，但在这方面理论研究的成熟和实际编制的完善化却经历了半个多世纪的漫长时间。早在三十年代印度著名的图书馆学家阮冈纳赞首先明确提出了分面组配的理论，并运用这个理论于1933年编制出第一部分面分类法——《冒号分类法》。在他的分面分析理论的影响下，英国分类法研究小组又经过大量理论和实践研究，在1955年正式向英国图书馆学会和联合国教科文组织提出了《需要以分面分类法作为一切情报检索方法的基础》的备忘录。1957年在英国多金召开第一次国际分类法会议上，一致同意该备忘录的建议，从此这一具有国际影响的分面理论得到了国际图书情报界的肯定，为情报检索语言的

研制和改造指明了根本方向。七年后,在丹麦的埃尔西诺尔召开了第二次国际分类法会议。在这次会议上,传统的体系分类法受到了根本性的冲击,会议对传统分类法概念作出了新的广义的解释:“分类法是一种在众多单个语义单位之间建立属种或其他联系的方法,而不论其固有的等级性如何,也不论其能否兼用传统的或机械的情报检索系统。”这说明传统分类法的内涵已被彻底改造过了,这一被沿用近一百年的列举式树形知识结构发生了前所未有的变化,标志着分类法发展到一个新的阶段。近一、二十年,由于电子计算机在图书情报系统中的广泛利用,使编制大规模的分类主题一体化的词表成为可能。1970年英国情报学家J·艾奇逊(J.Aitchson)编制的《分面叙词表》问世。从此分类主题一体化问题在实践上得到了解决。继这第一部分面叙词表之后,相继出现了大批类似的词表,受到国际图书情报界的高度重视和评价。1982年在西德澳格斯堡召开的第四次国际分类法会议认为,分面叙词表是“更一致和更完美的分类法与叙词表编制方法结合的产物”,从而国际公认了分面叙词表是分类法与主题法结合的成功典范。

在我国对分类主题一体化理论问题的研究工作起步较晚,是从60年代开始的。那时专家们大都集注于两种检索语言相异点的研究,比较它们各自的优缺点,对两者基本因素相互渗透的关系很少进行考察。1962年,著名图书馆学家刘国钧先生对分类法和主题法的相异点进行比较研究时,发现了两者可以取长补短,合璧为一的趋势,他鲜明地指出:“分类法和主题法各有所长,各有所短,合则双美,离则两伤”。他还进一步强调:“近来,分类法和主题法各向对方转变,是很值得注意的一种趋势。”这恐怕就是我国分类主题

一体化思想的开端。图书馆学专家杜定友先生也曾提议编制分类主词目录（或称分类主题目录），试图探索分类主题一体化的新路子。但在那“冲击一切的”年代里，这些有益的思想和建议统统地被冲击掉了。从70年代末以来，我国情报检索语言的研究进入了蓬勃发展时期，出现了大量的论著和译著，对分类法与主题法的概念逻辑基础、结构功能特性、语义语法手段和词汇规范原则等问题进行了系统论述，为我国编制一体化词表，提供了理论先导。与此同时，编制一体化词表的实践也取得了很大进展。1983年山西太原机械学院把《中国图书分类法》中的“TJ武器工业类”与《汉语主题词表》中的“武器军事技术装配”方面的叙词，改编为《常规武器工业分面主题词表》，这是我国图书情报界在编制专业分面叙词表方面，获得成功的首次尝试。1987年北京图书馆组织全国38个单位开始编制《中国图书分类主题词表》，这部综合性的对应表是从当前全国的实际情况出发，创制出来的一条实现分类主题一体化的捷径，对建立我国统一的情报检索体系必将起促进作用。1987年中国社会科学院文献情报中心为建立全国的社会科学数据库，采用分面叙词表的最新模式开始研究和编制《社会科学检索词表》。这些理论和实践方面的努力，预示我国一体化检索语言的发展将出现良好的前景。

为更深入地学习和借鉴国外在编制分面叙词表方面的理论、经验和技术，我们选择了九篇既有代表性又有实用性的论文和一篇综述国外分类主题一体化词表进展情况的文章，汇编成册，奉献给我国图书情报界的同行。

第一篇文章《统一的自动化检索系统中的主题法与分类法的相互关系》，选自1975年苏联的《国外图书馆学和目录学》杂志，为苏联学者维连斯卡娅撰写。此文较为全面

地评述了国外情报系统发展中分类、主题一体化的新趋势，从理论与实践两个方面对这一新趋势进行了较深入的研究，并从哲学宏观的认识论角度出发，正确的指出，主题法与分类法是认识世界的两种不同的方法，它们有着共同的认识论的基础。这篇文章能帮助我们了解国外分类、主题一体化理论与实践的总的发展概貌。

《十进分类法和叙词表概念的一体化》选自1978年召开的第39届国际文献工作联合会（FID）出版的《文献和情报的新趋势》集子中的一篇论文纲要。它是由西班牙学者拉·罗·德尔加多（R·R·Delgado）撰写的。此文用系统论解决分类主题一体化的理论和实践问题。作者指出，第二次世界大战后发展起来的系统论是把科学思维和术语结合成一个有意义的整体的重要尝试，它是试图把人类全部知识和活动纳入一个统一的概念体系的跨学科的方法。作者提出元语言（Metalanguage）的概念，并用以作统一学科发展的一种通用的文献语言，对社会科学和自然科学的概念和活动进行重新排序。他运用这种系统论方法，具体把人类的全部知识分成1—6级类，其中1—2级类是概念核心的不同方面；3—6级类表示各个方面的子系统，一共可容纳100万个语词。这样，任何深度的分类法均可包容进去。在这个大系统中还可以通过参照系统来沟通同级系统之间，系统与子系统之间，系统与周围环境之间，系统与其他不同级和不同类别的系统之间，概念、词语与范畴之间的种种网状关系。

在我国尚无运用系统论的方法解决分类主题一体化问题的尝试，这种突破旧有一切分类模式的新方法，无疑可以开阔眼界，广辟思路，在我们传统的思维方式中注入新鲜内容。

关于编制一体化词表的技术方法问题，我们选择了三篇文章。

《用分类法提供的资料编制叙词表——布利斯书目分类法作为叙词表的词汇和结构来源》选自1986年英国《文献杂志》第3期，作者是英国分类法小组成员著名的情报学家艾奇逊女士，她根据自己编制多部分面叙词表的丰富理论和实践经验，提出了以布利斯书目分类法第2版（以下简称BC2）为基础，来编制各门学科领域的叙词表的有益建议。BC2是由英国情报学家J.米尔斯（J. Mills）主持编写的，它是在第一版《书目分类法》（1940—1953）的框架基础上发展成的一个综合性的分面分类法。BC2保留了第一版的大类结构，而在每一学科大类里实行了分面列类，共分事物、事物性质、事物的部分、事物的组成因素、属性、过程、操作、手段、时间和空间等十一个组面。此表采取了一些技术措施，使分面结构既具有简明性，又具有最大限度的容纳性，同时还能与叙词表互相兼容。BC2从1977年开始陆续出版，到1985年已经按学科分卷出版了7册。待出版和正在编制的共约20余卷，从中可以获取各学科领域的分面结构和词汇来源。当今有许多分面叙词表，在词汇和词间关系方面都受益于它。

正因为本表产生于对分类法研究走在世界前列的英国，并且它的分面框架一直受到国际图书情报界的赞誉，所以及时收集不断出版的BC2，对我国编制各种专业叙词表具有重要的现实意义。

艾奇逊在本文具体讨论了以《BC2》作为叙词表的词汇和结构来源所存在的问题、利弊及前景。她强调指出，BC2是叙词表的词汇和结构来源，其价值应受到编制者的重视。

目前国外可以提供我们的现成社会科学分面分类法为数不多，BC2 有的学科如社会学和宗教等学科，甚至可以直接效仿。尽管此表代表西方国家的理论观点，在学科划分和词汇的内涵上，与我们不尽相同，但是它的分面组配框架、词汇参照系统、标识符号系统都是值得我们直接借鉴的依据。

《社会科学叙词表的综合》是英国情报学家艾奇逊提交 1981 年 5 月 24—27 日在西德比勒费尔德召开的“社会科学概念和术语分析 (CONTA) 会议”的论文。本文全面论述了 1980 年联合国教科文组织计划编制《社会科学综合叙词表》

(Social Science Integrated thesaurus) 的历史背景、目的、特点、步骤和方法。世界各国社会科学术语很不一致，对社会科学一些术语又各自赋予它不同的含义，从而使各国社会科学叙词表之间无法兼容，严重地妨碍了社会科学情报的交流。为此，联合国教科文组织决定汇编一部能包容目前所有社会科学情报语言的综合叙词表。

这种模式是跨国界、跨民族、跨语种的一个创举。它吸收了现有各国分类法和叙词表的编制成果，试图提供一个世界社会科学情报检索语言相互沟通的工具，无疑这是一项极有价值而又十分艰巨的工程。作者是本表的主持人。她仅就“失业”这一小主题领域的主要款目，就选用了五种情报语言，设计出矩阵显示方案，对每个款目逐一做了相应的模拟示范，因各国国情不同而出现的“解雇”、“免职”、“失业”等各种不同表现形式都能清晰地显示出来。艾奇逊指出，综合叙词表能够作为改进社会科学交流和情报索引的参考标准和交换工具，“它将促进现有情报语言的改善，增强对社会科学概念的全面理解。”国际科技情报检索网络化迅速发展，我国社会科学情报数据库进入国际网络只是时间问题。因此，

我们在研究和编制社会科学叙词表时，从现在起就应考虑国际上通用的标准，在分类框架结构、叙词规范等方面尽量与其靠拢，以利交流，做到国际情报资源共享。这也正是我们选择这篇文章的意图所在。

《自然语言索引系统中的分类原则》选自由 A.梅尔特比 (Arther Malthy) 编辑，1972年伦敦出版的《70年代的分类法》一书中的一章。由英国著名的情报学家 B.C. 维克里 (B.C.Vikery) 撰写。作者为把分类因素准确地引入字顺索引，设计出二道不可分割的基本工序：一是分组和划分，即把在某些方面相同的实体加以集中，而把那些在某方面不相同的实体予以分散；二是排序，即把前一种相同的实体按其意义顺序进行排列。为加强分类的显示功能，维克里对字顺索引和字顺词表中出现的多组词语、短语和复合主题之间的语义关系又提出了分组划分和排序的原则和方法，即倒置标题、副标题和相互参照三种分类方法以及必须遵循的倒置规则、范畴分析、关系分析和分面等级体系四项分类原则。

这些原则和方法是维克里在研究情报学家们不同理论观点的基础上提出来的，在国际图书情报界颇具影响。

为了使读者了解苏联在情报检索语言方面的进展，我们从苏联《情报理论与实践问题》(全苏科技情报所编，1985年出版)一书中选译了一章——《社会科学自动化情报系统的语言保障：组成与结构》。本章是由 P.P. 姆季瓦尼和 B.A. 格林斯基合著的。作者更深入具体地论述了苏联科学院社会科学学术情报所建立的苏联社会科学情报检索语言系统组成和结构问题。苏联社会科学情报检索语言独具特色，是有别于西方的“社会主义检索语言模式”。苏联是社会主义国家，

其意识形态是以马克思列宁主义为指导的，因此苏联社会科学的性质在其主导方面与西方国家有着本质的不同。作者在论述编制苏联社会科学情报检索语言的特点时，充分地注意到这一根本不同点。他着重指出，社会科学情报在政治上是不可能中立的，应当鲜明地判定“文献内容的社会本质、作者的思想立场、作者的流派属性等等”。在这方面，对我们不无参考价值。例如本文指出，马列主义科学的一系列基本概念，像“基础与上层建筑”范畴在西方国家那里找不到或者被他们作了歪曲的解释，而像“非意识形态化理论”，“政治多元论”、“欧洲共产主义”等叙词却被西方国家称之为新概念、新理论和新学派的术语，实质上是属于资产阶级的理论。另外对一些概念和问题的解释往往与西方国家存在不同的观点。

针对上述社会科学情报检索语言的政治思想性特点，苏联在编制一体化的叙词表过程中，“把主要的注意力集中在编制规范词汇表上。”并具体制定了一系列反映思想意识面的原则和方法。如规定凡是资产阶级学者应用的表示资本主义社会现象和资产阶级理论和观念的概念和术语可以收进叙词表，而煽动性术语只能有限制地收入表内，并加附注，给予评价。又如采用通用科学概念辅助分面表来判定文献作者的立场等等。

上述反映苏联情报检索语言研究现状的论文在一些方面对我们是很有参考价值的。因为我国的社会制度、社会科学研究理论、观点和方法与苏联有许多相同点或相似的地方。

我们还选择了三部国际公认的、有代表性的词表的导言，以便更加具体了解分面叙词表的新模式。

第一部是1970年首次成功问世的《分面叙词表》(Thesaurofacet)，由英国分类法小组成员艾奇逊女士主编。本表通过控制词汇及分面技术，将原来的英国电气公司的《电气工程分面分类法》第三版进行了全面修订。新编成的词表，把分面分类表和叙词表两部分结合为一体，使每个叙词分别出现在分类表和叙词表中，二部分之间用分类号互相联系。作者在导言中，着重强调了分面叙词表的形式和结构，分类表的作用，先组词与组配的关系以及词表中的组配问题。该词表问世后，受到国际图书情报界的高度重视，英、美、苏等国的情报学家相继撰文加以评价。例如，苏联维连斯卡娅认为，该分面叙词表的功绩在于“首次建立了一个真正实现了分类法和主题法两者合一的系统”，“拟定了方法，建造了模型，提出了在不同体系中协调和采用统一术语的具体措施。”

第二部是1977年由艾奇逊编制的另一部重要的分面叙词表《联合国教科文组织叙词表》(Unesco Thesaurus)。这部词表是为联合国教科文组织的计算机文献系统编制的。收录范围涉及哲学、社会科学、人文科学、自然科学和技术科学领域，是一部综合性叙词表。由于此表成功地运用了电子计算机的功能，使分类表自动地转换生成了字顺叙词表、族系表和轮排表三个组成部分，从而大大缩短了编表的时间，提高了词表的质量，真正做到了分类表的语词概念与字顺表的语词概念一一对应。作者在论述这些特点的同时，还着重对本表丰富的词汇来源，分面分析的结构特点以及计算机管理词表更新的办法进行了重点评述。

第三部是1981年由英国标准协会(BSI)组织并正式聘请艾奇逊担任顾问编制的《基础叙词表》(Root Thesaurus)。

该词表“集二十年来分类语言及描述语言成果之大成,也是十年来分类主题一体化的最新成果。”由于这部分词表进一步完善了分面叙词表的模式和编制技巧,受到了1982年在西德召开的第四次国际分类法研究会议的高度评价。我国国家标准局标准资料馆已将此表译成中文,它将为我国编制有关词表提供重要的术语来源。编写者熟练地运用了分面分类和词汇控制的技术,通过对词表宏观结构的简化与微观结构的精化办法,来增强词表的功能。拿微观结构与其他一体化词表相比,它的组成要素更丰富了,结构方式更趋合理了。导言中着重指出该表的特点是:(1)在宏观结构上已经明确以字顺主题显示表为主,以辅助性的字顺表为辅。而传统的词表一直是以字顺表为主,这是一个新的进展,更加突出了分面分析的应用问题。(2)首次运用国际通用的数学符号作为参照系统的标识符号,增强了直观性。(3)机编词表技术的应用功能得到更加广泛的发挥:既可以由主题显示表转换生成字顺表,又可以用于词表的管理系统,还可以利用主文档派生出各种专业词表及不同语种的词表。英国标准协会还计划建立第二文档,存放所有叙词的定义,以进行科技术语的标准化、管理。

这三篇导言,不仅为我们介绍了分面叙词表编制的最新成果,而且导言本身也是对一体化应用研究的范文。

最后一篇是侯汉清于1987年3月在中国社会科学院文献情报中心课题组召开的“关于分类主题一体化”理论讨论会上发表的论文《国外分类主题一体化词表的进展》。该文对国外一体化词表的四种模式:分面叙词表、叙词表式索引、分类表——叙词表对照索引及整体化叙词表的进展情况,进行了全面的论述和评估。为使读者较系统地了解国外编制分类主

题一体化词表的进展情况，亦收入本集。

上述十篇文章只是反映国外一体化情报检索语言的发展概貌，我们不可能、也没有必要将国际上有关的所有情况全部引进过来。因为任何成功的经验都有它一定的局限性，只能借鉴，不能照搬。我国对一体化情报检索语言的研究经历了初探、中断和发展的曲折道路，有深刻的教训，但近年来，也确实取得了明显的成绩。相比较而言，引进（现在引进还很不够）多于独创。而要取得我国一体化情报检索语言的大发展，必须立足本国实际，首先从理论上进行大胆地探索和创新，以指导实践的进一步发展。实现我国情报检索语言的一体化，存在着一系列的问题，我们必须发扬独创精神，探求这一新领域的特殊规律，要源于实践、指导实践，发展具有中国特色的一体化情报检索语言。

吕其苏

一九八八年春节

目 录

前言	吕其苏 (I)
统一自动检索系统中主题法与分类法的相互关系	
〔苏〕C. K. 维连斯卡娅著 张涵译 朱铁声校	
.....	(1)
十进分类法和叙词表概念的一体化 〔西班牙〕R.	
R. 德尔加多著 侯汉清、储荷婷译	(27)
用分类法提供的资料编制叙词表——用布利斯书目	
分类法作为叙词表的词汇和结构的来源 〔英〕J.	
艾奇逊著 吴政 侯汉清译	(45)
社会科学叙词表的综合 〔英〕J. 艾奇逊著 戴维民译	
侯汉清校	(87)
自然语言索引系统中的分类原则 〔英〕B. C. 维克	
里著 侯汉清 译	(124)
社会科学自动化情报系统的语言保障：组成与结构	
〔苏〕P. P. 姆季瓦尼, B. A. 格林斯基著	
朱铁声译	(155)
《分面叙词表》导言 〔英〕J. 艾奇逊著 侯汉清译	
.....	(181)
《联合国教科文组织叙词表》导言 〔英〕J. 艾奇逊著	
吕其苏译 侯汉清校	(204)
《基础叙词表》导言 侯汉清译	(237)
国外分类主题一体化词表的进展 侯汉清	(252)

统一自动检索系统中主题法与分类法的相互关系

——国外情报检索系统理论与实践的新趋势

〔苏〕C.K.维连斯卡娅 著

张涵 译 朱铁声 校

前 言

主题法和分类法的相互关系是当代分类理论中急待解决的问题之一。这是一方面因为人们对情报检索系统的要求不断增高，另一方面因为人们力图组织国际的情报交流。电子计算机在图书情报工作中的广泛使用和情报检索语言理论的发展为实际解决这个问题创造了条件。

本文阐述的是反映分类理论发展新趋势的某些主题法和分类法相互关系的问题，并介绍国外在建立“混合的”、一体化的情报检索系统方面的初步经验。

分类法与主题法是对文献进行分析—综合加工的最重要的方法。这与现有的情报提问和检索的形式——体系和主题形式——相适应。前者是为了按照总的知识门类、知识门类的各个部分、或该知识门类中的个别主题来提供文献，而后者则是为了按照问题、主题、题目提供综合性的、跨门类的文献。

我们已经谈过我们的观点（5）（括号中的数字代表本