

情报检索语言综论

戴维民 著



军事谊文出版社

情报检索语言丛书

情报检索语言综论

戴维民 著

军事谊文出版社

情报检索语言丛书

侯汉清 陈树年 戴维民 主编

(京)新登字121号

情报检索语言综论

戴维民著

*

军事谊文出版社

(北京安定门外黄寺大街乙一号,邮编100011)

850×1168毫米 32开本 15印张 400千字

1994年9月第一版 1994年9月第1次印刷

印数:1—1000册

ISBN 7-80027-5181-3/C·48

定价 18元

序 言

张琪玉

戴维民同志的《情报检索语言综论》一书收录了37篇论文和5个附录,这些文章涉及情报检索语言领域的许多方面,旁征博引,内容丰富,是学习和研究情报检索语言不可忽略的一部著作。

我们的情报检索语言理论研究工作必须开辟新路。要从四个方面着手,一是改变研究方向,二是扩大研究范围,三是改进研究方法,四是吸取国外的研究成果。*

我国在闭关锁国多年之后,借鉴国外的研究成果,对较快地缩短我国与国外在情报检索语言理论研究方面的差距,具有重要的意义。在引进国外研究成果方面,侯汉清同志和戴维民同志是工作做得最多的两位,作出了较大的贡献。侯汉清同志的《当代分类法主题法索引法研究》一书和戴维民同志的这本书都是通向国外的窗口,从这窗口里,我们可以获得国外情报检索语言领域的许多信息和新的知识。当然,“洋东西只是一种补品,吃下去,消化了化成自己的血肉,才能壮身子”。情报检索语言都具有或多或少的国家特点和自然语言特点。也就是说,要符合国情。我们固然要借鉴国外情报检索语言的原理和方法,乃至直接采用国外分类表、词表中可用的部分,但是必须明确,我们是为着创制中国的情报检索语言而去研究国外的成果,因此在学习外国的同时,还必须用主要精力来研究中国情报检索语言

的各种特殊问题,必须把学习外国与研究中国很好地结合起来,把学习和创新很好地结合起来。否则,我们将创制不出不但在原理和方法上具有世界先进水平,而且又是符合中国国情的情报检索语言来的。”就这方面来说,戴维民同志也做了很多工作。他的许多文章,总是对中外情况作比较研究,他对我国情报检索语言的实际问题也作了很多研究,这在他的书里也是可明显看出的。

戴维民同志是一位青年学者,他在华东师范大学读书期间,正是80年代我国情报检索语言和情报语言学的新的发展时期,他以执着的精神,参与了对情报检索语言的求索。11年的积累,结出了硕果。我更希望他今后能取得更多的成果。

1994.7.8

张琪玉:论情报检索语言的研究、创制与普及。载:《图书情报知识》1983年第4期

· I ·

自序

情报检索语言作为一种实体尽管它源远流长,但作为一种研究对象,情报语言学作为一门新兴学科,历史并不长,真正的发展和深入研究则是五六十年代以后。而我国起步则更晚,虽然五六十年代也曾有过一段发展历史,但到了八十年代才进入了比较繁荣时期,形成了学科体系,有了学科带头人和学科的代表著作,实践上更是得到了很大的发展。情报语言学已成为图书馆学和情报学的一个比较成熟的分支学科。

随着现代信息技术的飞速发展,情报检索技术也在发生巨大变化,作为情报检索的语言保证的情报检索语言必须适应变化,不断创新。从发展趋势看,情报语言学应着力研究自然语言在情报检索中应用问题,这并不意味着受控的情报检索语言将被放弃使用,而是要如何把情报检索语言的基本原理和有效手段应用到自然语言情报检索中去。要降低词汇控制和标引的难度和成本,同时也要力图提高情报检索的效率,这是当今计算机情报检索时代重要的研究课题,这一问题的重大突破,在未来高速信息网的建成后,仍能广泛地适应信息检索的需要。挑战与机遇并存,传统的研究内容要不断更新,现行的探索领域要不断拓展,情报检索语言这一领域需要有更多的理论家和实践者,愿有更多的有志者加入这一研究领域。

当我80年代初进入大学学习时,就开始了对情报检索语言某些问题的初步探索,也激发了我进一步钻研的欲望。于是我研究生的研究方向也选择了情报检索语言这一领域。三年的研读,从导师何金铨和刘重燕先生处获益非浅,情报检索语言界的前辈张琪玉、李兴辉、刘湘生、侯汉清、丘峰、陈树年等专家学者也给予了我积极的引导和扶携,使我在情报语言学这一园地有志耕耘了十多年,本书就是我这

些年从事情报语言学研究的一点总结,虽然还不太成熟,却描绘了我学术生涯的一点轨迹。

本书选收了我十多年来关于情报检索语言研究的37篇文章和5个附录,有些内容因理论与实践新发展有所变化,为了保持原貌,未作大的修改,只作了一些文字上校改。在侯汉清教授鼓励下,集结出版,权且作为引玉之砖。张琪玉教授拨冗为本书作序,并给予我过高的评价和真诚的勉励。在本书出版之际,谨向给予指导、扶携的学术前辈们,为本书出版给予大力帮助的其他许多同志表示谢意。

戴维民

1994年教师节于沪上

目 录

序言

自序

第一篇 情报检索语言总论	(1)
情报检索语言及其分类	(2)
论我国情报检索语言研究道路	(8)
论情报检索语言的文献保证原则	(20)
情报检索语言句法结构简论	(32)
情报检索语言的组配及其句法语义控制	(43)
引用次序研究	(50)
论轮排技术	(71)
情报检索语言句法控制符号概述	(83)
英国分类法小组及其对情报检索语言的贡献	(92)
我国分类主题一体化研究进展	(101)
第二篇 分类检索语言研究	(115)
简论培根的知识分类及其对图书分类法的影响	(116)
知识的结构与图书分类法的未来	(123)
中国现代图书分类法	(130)
综合科学的崛起与图书分类体系的变革	(134)
《中图法》作为我国标准分类法的几个问题	(150)
我国专业分类法编制浅析	(158)
附:土壤科学分类表	(167)
食品工艺分类表	(176)
知识组织文献分类法	(178)
信息编码分类法	(193)
《中国图书资料分类法》的发展历程与路向	(195)

《中国图书资料分类法》分面改造研究.....	(202)
《中国图书分类法》在台湾省的发展和应用.....	(240)
分类标记.....	(244)
分类目录的现状与未来评析.....	(248)
外国图书分类法.....	(256)
国外图书分类理论发展述要.....	(261)
国外图书分类法修订的理论与实践.....	(272)
《布利斯书目分类法》第二版评介.....	(282)
附:《布利斯书目分类法》第二版“J 教育”类导言	(296)
第三篇 主题检索语言研究	(307)
叙词语言的发展及原理.....	(308)
汉语叙词表的发展现状与未.....	(317)
简论专有叙词.....	(325)
叙词表评价论略.....	(330)
叙词表传统模式的突破——《教育分面叙词表》的编制理论 与实践.....	(338)
《汉语主题词表》及其评价.....	(355)
《军用主题词表》评介.....	(362)
主题法标准化.....	(374)
计算机在叙词表编制中的应用.....	(380)
第四篇 标引与检索问题研究	(386)
标引一致性述略.....	(387)
自然语言标引与检索现状与趋势.....	(404)
关键词索引的类型及发展.....	(411)

第一篇 情报检索语言总论

- 情报检索语言及其分类
- 论我国情报检索语言研究道路
- 论情报检索语言的文献保证原则
- 情报检索语言句法结构简论
- 情报检索语言的组配及其句法语义控制
- 引用次序研究
- 论轮排技术
- 情报检索语言句法控制符号概述
- 英国分类法小组及其对情报检索语言的贡献
- 我国分类主题一体化研究进展

情报检索语言及其分类

情报检索语言是情报检索系统的重要组成部分,是情报检索的语言保证。无论是情报存贮或是检索都离不开检索语言。因此,掌握情报检索语言是文献检索的重要环节。

一、什么是情报检索语言

情报检索语言在国外则称之为标引语言或索引语言。

什么是情报检索语言呢?张珙玉教授的定义是“根据情报检索的需要而创制的人工语言”。苏联著名情报学家切尔内对情报检索语言则作如下阐述:“一种专门的人工语言,它用于描述(表示)文献的中心问题或主题和外表特征,以便以后在其它文献集合中查找需要的文献,或者用于表示情报提问的内容和检索的需要的文献。”

理想的情报检索语言自身必须具有词汇和语法手段,能准确有效地用于情报标引和检索,真正起到情报检索的语言保证作用。

二、情报检索语言的类型

我们通常把一部分类法、一部主题词表称为一种情报检索语言,就目前全世界已有的情报检索语言来看,其数量已有几千种之多。为了对每种检索语言都能准确地认识其结构和作用,我们对它们进行类型划分是非常必要的,而实际上对情报检索语言研究显然也是离不开对其进行分类的。下面是两种具有代表性的划分方法:

1. 张琪玉关于检索语言类型的划分:

分 类 标 准	语 言 类 型
按构成原理分	① 分类检索语言 ② 主题检索语言 ③ 代码检索语言
	④ 引证关系追溯法
按所包括的学科 或专业范围分	① 综合性语言 ② 专业性语言
按适用范围分	① 一种检索工具或检索系统专用语言 ② 某一类型图书馆或情报机构专用语言 ③ 各类型图书馆和情报机构通用语言 ④ 国际通用语言
	⑤ 某一类型文献资料专用语言
	⑥ 传统检索工具适用语言 ⑦ 机器检索系统适用语言
按采用的自然 语言分	① 单语种语言 ② 多语种语言
	③ 基于英语的检索语言 ④ 基于汉语的检索语言
按标识的组合 方式分	① 先组式语言 ② 后组式语言
	③ 先组散组式语言

2. 切尔内关于检索语言类型的划分:

1 先组情报检索语言

1.1 枚举式分类法

1.1.1 等级分类法

1.1.2 字顺主题分类法

1.2 组面分类法

2 后组情报检索语言

2.1 叙词语言

2.2 语义代码

2.3 结构段语言

3 题录引文语言

从上述两种划分看,由于认识情报检索语言的侧重点不同,也就是划分的标准不同就得出不同的结果,下面我们介绍一种情报检索语言的谱系式分类方法。所谓情报检索语言谱系分类,是指按各种语言的亲属关系的分类方法。谱系分类,也叫发生学分类。

在情报检索语言的形成、发展过程中,语言之间既相互渗透、相互综合,同时又不断分化,因此各种语言之间既有联系又有区别。采用谱系分类不仅为我们研究各种语言提供一条新的途径,同时可以看出情报检索语言的发展过程及其未来的发展趋势。

谱系分类是将情报检索语言按语系、语族、语支和语种体系划分的一种方法。

(1)语系。是谱系分类最上位概念。情报检索语言有两大语系,即文献内部特征检索语言和文献外部特征检索语言。

(2)语族。是语系的进一步划分,如文献内部特征检索语言又可分为分类语言、主题词语言等。

(3)语支。是语族的进一步划分,如分类语言可分为等级体系语言、分析—综合语言、半分析—综合语言等;主题词语言可划分为标
题词语言、单元词语言、叙词语言等。

(4)语种。指具体的以某一种语言词典(分类表、主题词表)为基础的语。如:《汉语主题词表》、《中国图书馆图书分类法》、《军事图书资料分类法》等。

除此之外,我们还将专门适应于某种文献类型的语言称之为不同的语体。如:图书文献分类法、档案文献分类法、专利文献分类法、期刊文献分类法等都是不同的语体。

在各种语言之中常用的是分类语言和主题词语言。

三、分类语言的语支

分类检索语言习惯简称分类语言或分类法,是具有悠久历史和广泛使用的检索语言类型。

分类语言主要有三个语支,即等级体系分类语言、分析—综合分类语言、半分析—综合分类语言。

1. 等级体系分类语言。这是一种直接体现知识分类的等级制概念标识系统,是对表示文献内容及某些外表特征的概念进行逻辑分类、系统排列而成的,它用等级体系显示类目的从属关系、并列关系等各种关系,以尽量列举来显示概念划分的完整性和专指性,它适合于从学科、专业系统检索和利用文献。我国目前使用的《中国图书馆图书分类法》、《军事图书资料分类法》、《中国科学院图书馆图书分类法》等就属此种类型的语言。

2. 分析—综合分类语言。这是以印度阮冈纳赞的《冒号分类法》为代表的语言类型。等级体系分类语言的一个共同特点是企图详尽无遗地列出已知的主题,并将每个主题配以固定的类号,但它们的缺点就是不可能将现有的主题详尽无遗地列举出来,更无法预见未来出现的新问题,而且还严重地受标记符号的限制,难以扩充增加新类目。因此,采取分析兼综合的方法,即先把一切主题分析成它的组成因素,编成表,并给每个因素配一个号码,然后再根据需要将有关的号码组成一个完整主题的标记符号,这就较好地解决了等级体系分

类语言的问题。

3. 半分析—综合分类语言。这是由上述两种类型分类语言相互结合而成的一种分类语言。有人认为《国际十进分类法》便属这类分类语言。

四、主题词语言的语支

主题检索语言,我们习惯简称主题词语言,又称主题词描述语言或主题法。主题检索语言发展很快,其中的叙词语言被认为是当代情报检索语言的主流之一。

主题词语言主要有三个语支,即标题词语言、单元词语言、叙词语言,有的也将关键词语言算作主题词语言的语支之一。其中叙词语言是现代应用最广的主题词语言,以至有时将主题词语言狭义地理解为叙词语言。

1. 标题词语言。标题词语言是早期主题词语言类型。它是以“标题”(Subject Heading)的语言形式表达文献的内容主题,并将其用作著录标目,按标题字顺排列的一种组织检索工具的检索语言。标题词语言最初见于1856年英国克里斯塔多罗的《图书馆的编目技术》一书,而正式创立则是美国图书馆学先驱克特。1876年,克特发表了《字典式目录条例》,其中运用的语言原理就是标题词语言。美国图书馆协会1895年印发了第一个权威标题表——《美国图书馆协会标题表》。1911年,《国会图书馆标题表》第1版出版,从而在相当长的一段时期里;标题词语言得到极大的发展。现在,美国的《国会图书馆标题表》已出至第10版,具有很大的权威性,影响全世界。但从标题词语言总体来说,却被后来新型的主题词语言逐渐取代。

2. 单元词语言。单元词语言与标题词语言相比,它的最大的不同就在于它是一种后组式的语言形式。最早由陶布提出,他称之为单元词组配索引法。陶布创造和设计了一种具有比号组配和反记著录特点的单元词卡片,这种卡片著录的内容包括两项:一是以每一个单元

词作为著录标目；二是在单元词标目下著录论及这一单元词内容的文献号码。在检索时，根据课题分析出构成这个课题的各个单元词因素，然后从单元词字顺目录中找出这些单元词卡片，用比号的方法找出这些单元词卡片上号码相同者，即为所要查找的文献号码，通过按图索骥，可最后找到所需文献。单元词语言曾一度风行，但由于在表达主题时语义不定性，所以很快被叙词语言所取代。

3. 叙词语言。叙词语言产生的历史并不长，自1947年到1950年美国穆尔斯创设“叙词法”以来，也只不过40年时间，但由于叙词语言综合了分类语言、标题词语言、单元词语言各自优点，而提供了一种科学的语言保证形式，使其发展很快。目前，世界各国拥有的叙词表的数量数以千计。叙词语言以概念组配为基础，具有对计算机检索系统的良好适应性。我国从70年代开始引进叙词语言以来，目前已拥有六、七十部叙词表，而且还有一批叙词表正在编制之中。可以预料，我国情报检索语言将全面进入叙词语言时代。

论我国情报检索语言研究道路

情报检索语言是情报检索的“语言保证”。作为“一门新兴的实验学科”^[1]的情报检索的发展,无疑将对情报检索语言不断提出新的要求。然而从我国情报检索语言理论和实践研究现状看,远远不能满足这一要求。从理论研究的视角来看,显然存在下列问题:把情报检索语言作为一种综合的语言实体进行研究少;对情报检索语言的理论基础缺乏研究;研究方法不健全,缺乏科学的方法论体系;对国外情报检索语言的最新成果研究吸收不够。面临这些问题,应选择一条能加快发展我国情报检索语言理论和实践的道路。笔者认为应从下面五个方面进行考虑。

一、把握情报检索语言的发展方向

近代情报检索语言权且以《杜威十进分类法》为开始,也有一个多世纪的历史了。印度著名图书馆学家阮冈纳赞曾将1876年《杜威十进分类法》第一版问世后的一百年称之图书分类法第一世纪^[2]。确实,在这一个多世纪里,情报检索语言经过不断发展,已初步形成了包括多语种多语体的语言体系,并且正朝着综合方向发展。

纵观情报检索语言的整个发展过程,可以看出下列几个发展方向:

1. 分面分析原理是所有情报检索语言的基础。尽管在《杜威十进分类法》、《国际十进分类法》、布朗的《主题分类法》以及奥特勒等人的论文中已有分面分析的最初思想,但是作为分面分析时代的开始和分面分析系统思想的建立都应归功于阮冈纳赞。阮冈纳赞在《冒号