情报检索词汇规范化

〔美〕F. W. 兰开斯特 著

QINGBAO JIANSUO CIHUI GUIFANHUA

科学技术文献出版社

情报检索词汇规范化

〔美〕F. W. 兰开斯特 著

杨劲夫 邱祖斌 孙荣科
赵立新 丘 峰 高维浦 译
温瑞林 杨鹤梅 匡奕祥

孙荣科 温瑞林 校

科学技术文献出版社

内 容 简 介

本书对世界各大情报系统使用的检索语言，从高度组织的规范化的标题法、叙词法和分类法，到非规范化的自然语言，都作了比较深入的分析和比较。介绍了美国为改进检索系统的查全率和查准率，提高其经济效益而作的各种尝试（包括成功的和失败的）。

全书共分二十五章，从规范化的角度出发，介绍检索语言的类型、结构和特点；词汇的选择、组织、展示、增补和更新；兼容性和互换性等问题。研究的重点是计算机检索用的叙词表和其他检索语言的规范化问题。书后附作者简介和主要检索语言简介等。

本书可供图书馆专业的师生阅读；也可供广大社会科学和自然科学的文献工作者使用。

VOCABULARY CONTROL
FOR INFORMATION RETRIEVAI
F. W. LANCASTER
INFORMATION RESOURCES PRESS
WASHINGTON, D. C. 1972

情报检索词汇规范化
〔美〕F. W. 兰开斯特 著
杨劲夫 等 译
孙荣科 等 校
科学技术文献出版社出版
北京印刷三厂印刷
新华书店北京发行所发行 各地新华书店经售
*
开本：787×1092 1/16 印张：16.5 字数：419千字
1982年12月北京第一版第一次印刷
印数：1—10,490册
科技新书目：36—62
统一书号：17176·353 定价：2.05元

译 者 的 话

本书是科技情报界的老前辈杨劲夫同志首先倡导翻译的。他译了前五章后因病搁笔，由其他同志继续完成本书的翻译。

结合前几年编辑和使用《航空科技资料主题表》，以及参加编辑《汉语主题词表》的工作体会，使我们感到如能在着手该项工作之前有机会阅读此书，也许我们的工作会是另一种面貌。这就是说，我们未能充分利用国外的这项工作经验。对搞科技情报的人未能利用情报中的情报来说，这是一件遗憾的事情。我们就是怀着这样一种心情来翻译此书，以作为我国同行共同改进我们工作的参考。

由于我们对本书介绍的某些内容还缺乏实际经验，理论上了解得也少，以及受语文水平的限制，我们的译文一定会有许多不足之处，请读者批评指正。

目　　录

第一章　词汇为什么要规范化？

情报检索*的一切复杂处理过程总要牵涉到类目——文献的这种或那种分类的处理。我们按主题内容标引文献时就是把它归入一个或几个类里，如图1。为了便于文献分类工作并进而熟练处理这些类目，每一类必须有一个类名或类标。给这些类目起的类名（我们也可把它们叫作“类标”）通常叫做标引词，而这些标引词的全体则叫做标引语言。为了满足一个特定课题所需的某种情报而在一种情报检索体系中进行查找时，我们所要做的是：

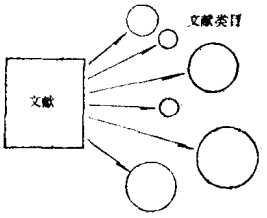

图1　经过标引形成的文献类目

1．判定哪些类目里最有可能有与所需情报有关的文献；

2．查找这些类目；

3．检出特定文献或所有文献。

一个情报检索系统的效率主要取决于该系统的类目规模和类目结构，同时也取决于准备查找的类目（即准备采用的查找方案）。一般来说，如果我们准备查找的都是大类，就能查全某一个特定课题的全部文献(即做到查全率高)，但却难于检出特定的文献(即查准率低)。反之，如果我们准备查找的是许多小类，就可能做到较高的查准率，但我们却会发现难于进行广泛查找而做到较高的查全率。

对于以后检索具有重要意义的是，在标引时，文献归类要遵循前后一致的原则。分类表要能起作用，它就必须把有关的文献归到一起。这就是说，我们必须使标引工作规范化。大多数检索系统都要求标引人员按事先的规定对文献进行归类，而不是给每个标引人员以全权而随意给文献设立新的类目，因为那样做会产生许多交叉类目，而使有关文献分散。这种事先的归定就是给标引人员编出一个标引时必须使用的标引词汇。标引人员判定手中的文献与什么问题有关，大体上对于哪些检索提问能提供有用的情报，就从正式词表里选出标引词给文献标上有关类目。例如，某一文献可能分为下列各类：DIET（饮食），CALCIOM(钙)，OSTEOPOROSIS（骨脆症）和RATS（鼠）。

规范化的标引词通常叫做规范化词汇或权威词汇。规范化词汇是一种粗略类型的标引语言。以后还要谈到，在一定情况下，可以使用一种根本不需要这种规范化词汇的检索系统。

词汇规范化的趋向是改善标引的前后一致性。两个标引人员（或同一标引人员在不同的时候）在表述一个特定课题的时候，比较可能选用较为互相一致的一个词或几个词，如果这些词是从事先编妥的表中选出来的，而不是在标引时独自编造的话。加之，当我们用一个检索系统进行查找的时候，如果我们能从一个确定了的类目表里查找，就较为可能查到正确的类目（也就是包含了我们关心的文献类目）。

规范化词汇在一个情报系统里的重要性见图2。它在标引过程和查找过程中都很突出。但规范化词汇不能，至少也不应该影响对文献进行的概念分析和对检索提问进行的概念分

---

* 本书通篇都用这个词，它通常用来表述本书所论述的检索活动和检索系统，即使所论体系不能检出情报本身（一种抽象概念），而只能检出一次文献或二次文献。

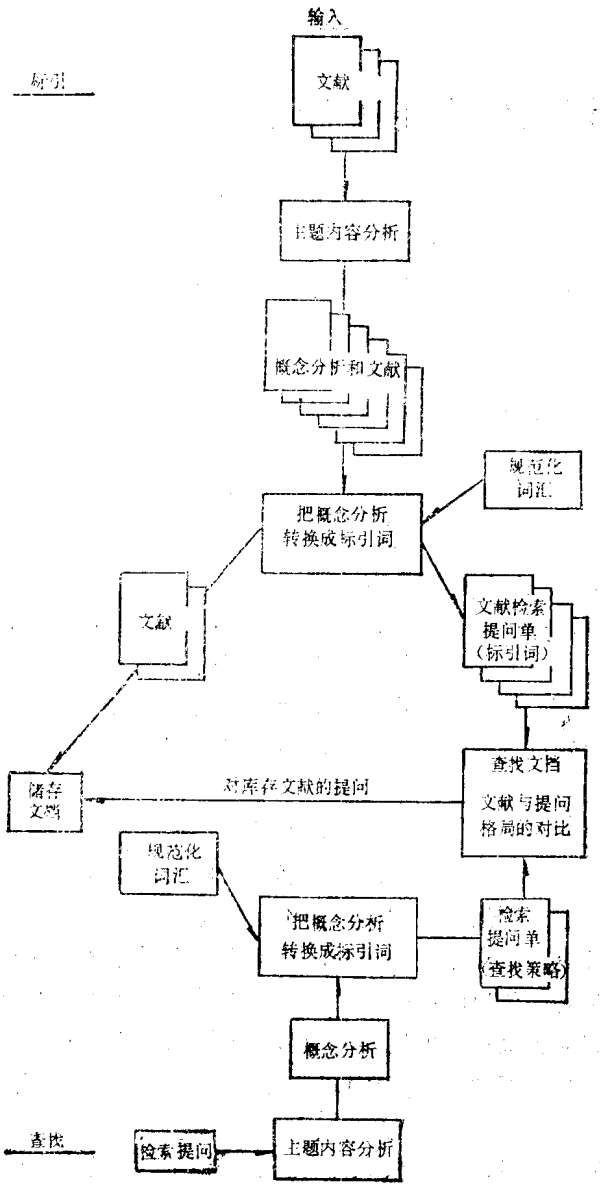

图 2　情报检索的输入输出过程

析。概念分析阶段同语言转换阶段是两个不同的阶段。我们首先要判定一篇文献或一个检索提问是关于干什么的，然后把我们的概念分析转换成标引语言中的词。概念分析和语言转换两个阶段对检索系统起着不同的作用。例如，我们能确切地判定一篇特定文献是涉及“氩弧焊”的。这就是我们的概念分析。当我们把它转换成标引语言的时候，我们可能要用含义更广泛的词（较不精确的词）来表述。检索系统的词汇不足以使我们精确定出涉及“氩弧焊”文献的类目。于是我们必须求助于含义较为广泛的词，例如“保护弧焊”，或“弧焊”，或“焊接”那样的词。

图 3 所示的是检索作业的各种步骤，并在旁边列出了对检索效能产生较大影响的每个因素。标引语言在以下两点上影响着检索效能：由于它确定了查找者能以多大的精确程度表述检索提问者感兴趣的事，它影响着查找策略；由于它确定了标引者能以多大的精确程度表述文献内容，它影响着标引作业。标引语言在检索工作中起着很重要作用，对检索系统全部任务的完成有着重要影响。

检索语言的存在主要是为了使标引者的语言与查找者的语言趋向一致。正常情况下，它对同义词和近同义词给予控制，以防止不同标引者对同一内容用不同的词来表述。词汇里指出优选的同义词，这就防止了同类文献的分散，并告诉查找者要查什么词，不要查什么词。正常情况下，它也把同形异义词区分开，如“plant”这个字就有“植物”的和“工厂”的以及其他各种可能的含义。

标引语言也有使我们易于进行族性检索的功用。它需以某种形式把相关的词归到一起，以便对较广泛的主题进行查找。假定我们需要查找“甾族化合物”（steroids）这个主题，词表就应该能够把以某种方法联结到一起的全部有关词展示出来。这就节约了查找者的精力。不然的话，他就得想出词汇里可能有的所有“甾族化合物”的词。这样，减少了他对相关词漏查的可能性，保证他实际上的查找是全面的。为了有助于标引者和查找者按上述那样，把

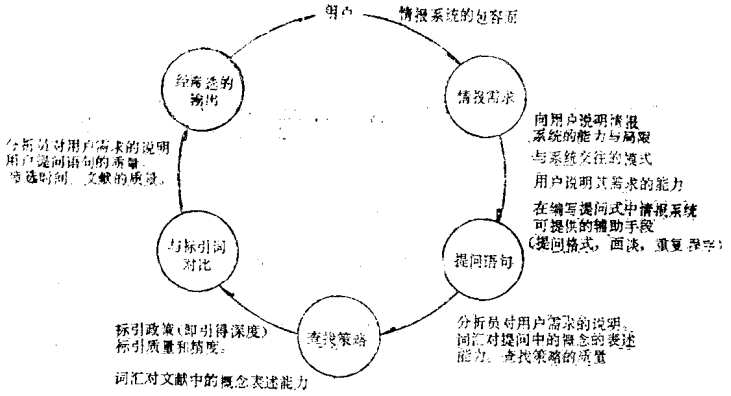

图 3　检索过程的每个步骤和每一步上影响检索效率的因素

词的关系展示出来，它会超出通常的属一种关系而扩大到其他类型的关系，包括部分-整体关系以及物质或工具同它的可能应用的关系。

（杨劲夫译）

第二章　词汇类型：先组式和后组式，枚举式和合成式

文献的主题内容往往是综合的，现代的情报检索系统必须能够表达任何程度的综合性。考虑一下一篇论述飞机、发动机和噪音的特定文献。标引这份文献时，我们也许想把它分到所有这三组概念类目里去，因为认为它同关于飞机、发动机和噪音的检索提问都有潜在的关系。但是，这份文献是同“飞机发动机噪音”这一特定的综合课题最有关系的，所以，我们必须能够标出这是“关于飞机发动机噪音诸文献”类目中的一份文献，因为检索系统应该能够针对这个课题的检索提问恰好检出这份文献。

基本上，我们可以有两种使用检索词汇的办法来标引和检出论述这种综合课题的文献。一种办法是定立一个唯一辨认这个特定类目的标引词。例如我们如果采用了“飞机发动机噪音”这个标引词，我们就是使用了一个表示“飞机”、“发动机”和“噪音”三个类目之间关系的词（图 4）。换言之，检索系统的词汇含有一个标志，它所辨认的类目是该表中存在的其他三个（文献）类目的逻辑积。这种词汇通常叫做先组式词；“飞机发动机噪音”是“飞机”、“发动机”和“噪音”三个词的事先组配的结果。用这种检索系统查找飞机上的发动机的噪音文献时，我们就看那个标着“飞机发动机噪音”的类目，期待这个类目里会容纳着与所查课题有关的文献。

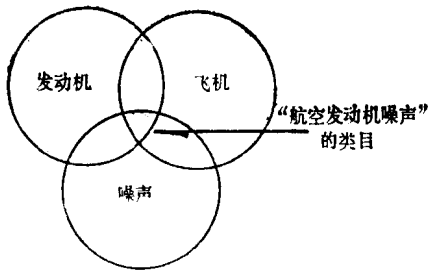

图 4　表示三个概念类目的交叉的先组式标引词

在另一个极端，是大部分事先不组配的各系统。这种检索系统的词汇里只定下一些比较基本的类目。最显著的例子就是陶布，M．(Taube, M.）编的单元词系统[3]。那个系统只让标引者使用一个个孤立的标引词（至少这系统的初版如此）。如对“飞机发动机噪音”这个课题就给文献标上不相关的标引词“飞机”、“发动机”和“噪音”。用这种检索系统进行查找时，我们得用某种方法处理文献的类目，以便认出三类都有而又大概会涉及“飞机发动机噪音”这一具体课题的文献。这种检索系统通常叫做后组式系统。后组式词能让我们于查找时调配它们的类目，以便得出它们的逻辑和、逻辑积和余项*。完全的先组式系统不能提供查找时调配文献类目的方便条件。这种系统里，查找工作受着标引语言中编定了的类间关系的限制。也就是因为这个缘故，贝尼埃[2]把后组式系统称为调配得了的，先组式系统是调配不了的。

后组式检索系统的存在不到三十年。最早发展的有巴顿的光重合原理的比孔卡[1]，穆尔斯的边缘穿孔卡[6]和陶布的单元词卡[4]。兰开斯特从前出的一本书里对这些系统有简要的叙述[5]。有些小的检索系统还在采用人工或机械编类。但许多小检索系统和大多数大检索系统现在都更充分地机械化了，用数字计算机以联机操作或脱机操作的方式运行，来调配和查找

* 文件各类之间正常关系可用布尔代数予以限定。用布尔代数编成查找策略的作法见兰开斯特著作[5]和 Principles of MEDLARS[7]。

文献的类目。其他检索系统则用缩微胶卷来达到同样的目的。

1940年之前，检索系统基本上是先组式，而且是调配不了的，最常见的是卡片式目录或印刷的索引。这样检索系统按字顺编排就叫做字顺主题目录或字顺主题索引*。若按某种分类体制的次序编排，这种检索系统就叫做分类目录或分类索引。字顺主题目录在美国一直为大众喜用，而分类目录在欧洲更为流行。字典式目录只不过是著录有作者和篇名的，按相同字顺编排的字顺主题目录。

先组系统和后组系统的差别是一个重大问题，而枚举式词汇和合成式词汇之间的差别也是重要的。按定义，后组式系统是合成式的，而先组式系统可以是完全枚举式的，也可是部分枚举式、部分合成式的。让我们看一看用来标引文献的两种不同的词汇，也就是列着标引词的单子。一个词汇的使用规则只许我们单独地使用各个标引词，而不许把它们结合起来表示综合一些的东西。这样的词汇是道地的枚举式。它列出或枚举了标引者在表述文献内容时必须使用的标引词，没有可以使标引者将表中所列的词结合起来（即合成）创立新标引词的变通余地。另一方面，第二种词汇，不但列有标引者标引时可用的词，而且又有可以将这些词以各种方式结合成新的，更专指的词的使用规则。这种词汇就是合成式词汇。

枚举式系统对于论述“飞机发动机噪音”的文献会要求将它标在“飞机发动机”之下，再另标在“噪音”之下。我们没有办法把这三个词弄到一起立个新词：“飞机发动机噪音”。在合成词汇里，我们可以把已有的词结合成新的、更专指的词。这样，在合成体制里我们就能够把“飞机发动机”一词和“噪音”一词合在一起，立一个能精确定出“飞机发动机噪音”这个类目的新类标。明显的是，如果我们用道地的枚举式词汇，表述文献内容能达到的专指程度受到编表人所提供的标引词专指程度的限制；在合成词汇里，这些限制就不存在了，我们多多少少有随意创立新的、专指的标引词的自由。有的分类体制（如国会图书馆的）几乎是完全的枚举式，而有些（如阮冈纳赞的《冒号分类法》）允许有相当数量的合成。标题表(subject heading)主要是枚举式，但允许某些合成（如主标题与子标题结合）。

从概念上说，先组式系统同后组式系统之间有明显的差别。但如果我们认真追查到语言的水平上，这差别就不那么明显了。例如，英语里有些字“概念上就是先组式”，因为它们表示的是两个或更多概念之间的关系。例如“尿分析(Urinalysis)”这个字的含义就是“分析尿(analysis of the urine)”，所表示的就是“尿(Urine)”这个课题和“分析(Analysis)”这个课题之间的关系（见图5）。同样，“蛋白尿(Protein uria)”这个字指的是尿里有蛋白，而“白蛋白尿（Albuminuria)”这个字指的是尿里有白蛋白（Albumin)。可是没有表示“尿里有结石（Calculi in urine)”这个意思的单字。我们就得把论述这个课题的文献既标引在“结石（CALCULI)”这个词之下，又标在“URINE(尿)”这个词之下，或者使用能结合这两个词的合成办法；例如把其中之一做为子标题（如：“结石，尿（CALCULI，URINE)”或“尿，结石（URINE，CALCULI)”），或立一个由词组构成的标引词，即“尿里的结石(CALCULI IN URINE)”，或“尿的结石（URINARY CALCULI)”。这些是英语的语格（accident)。它们使得情报检索用的词汇规范化问题麻烦

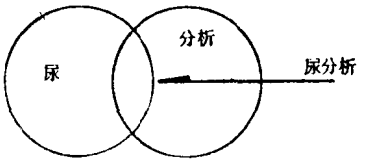

图5 “概念上先组”词的示例

* “目录”一词常限于指那种收录整本著作（书、报告等）的工具书，而“索引”一词是指收录著作某一部分（如期刊上的文章）的工具书。

起来。别的语言里，一个单字表示概念之间复杂关系的这种麻烦程度变化不一。举一个例，如德语就可以做到高度的概念先组。刘易斯·卡罗尔(Lewis Carroll)的混成字就是把概念结合在一个单字里的恰当例子(如：Slithy)。詹母斯·乔伊斯(James Joyce)杜撰了许多属于这种类型的另外的字，有些合并字在日常语言里使用，如："Smog 烟雾"("Smoke 烟"同"Fog 雾"的合并字)，"Mo-tel汽车游客旅馆"("Motorist 汽车游客"同"Hotel 旅馆"的合并字)和"Guesstimate，瞎估计"("Guess 猜测"同"Estimate 估算"的合并字)。

许多表面看起来是单纯的英文单字，实际上表示的并不是单纯的概念，而是可以分解为组元或因素(component aspects or factors)。例如"Thermometer"一字，表示的是计量温度的仪器(an instrument for measurement of temperature)(图6)。把字分解为其组元和基本含义的工作叫做"词义分解"。后面还要更详细地讨论它。

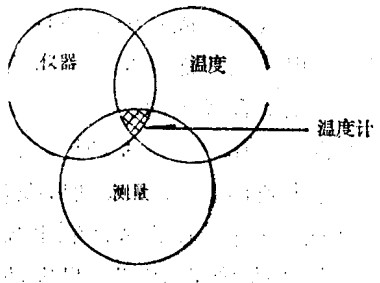

图3　词义分解示例

最现代化的检索系统是部分先组、部分后组的。常常一起出现的概念可以结合成为先组标引词，这个词于查找时也可以和其他词组配。例如，国家医学图书馆用的医学文献分析与检索系统(MEDLARS)里的词汇《医学标题表》就收有"肝瘤(LIVER NEOPLASMS)"这个先组词。如果要查"肝肿瘤(Liver Tumor)"的射线疗法，查找文献时，就可把这个词同"射线疗法(RADIOTHERAPY)"组配。

参　考　资　料

[1] Batten, W. E. "A Punched Card System of Indexing to Meet Special Requirements." Report of the 22nd ASLIB Conference, 1947, 37—39.

[2] Bernier, C. L. "Correlative Indexes: 1 Alphabetical Correlative Indexes." American Documentation 7 (1956): 283—288.

[3] Documentation Incorporated. Installation Manual for the Uniterm System of Coordinate Indexing. Dayton, Ohio: Document Service Center, 1953.

[4] Jaster, J. J. et al. The State of the Art of Coordinate Indexing. Washington, D. C.: Documentation Incorporated, 1962.

[5] Lancaster, F. W. Information Retrieval Systems: Characteristics, Testing, and Evaluation. New York: Wiley, 1968.

[6] Mooers, C. N. "Zatocoding Applied to Mechanical Organization of Knowledge." American Documentation 2. (1951): 20—32.

[7] National Library of Medicine. Principles of MEDLARS. Bethesda, Md. 1970.

(杨劲夫译)

第三章　分类法的词汇规范化

1876年出现的杜威，M.[5]的《十进分类法》是用分类法表述文献内容的首次尝试。原先的打算是给人们一个实用的图书馆图书排架手段，这个系统也用来（特别是英国）排列分类目录。这种分类法主要是枚举式；也就是它提供的是辨认各种文献类目的（既有单元类目也有复合类目）标志的表。在这种分类法的一切系统里，给文献的实际标志是字母的或数字的代码（分类标记）。这种标记代表了分类法编者赋予各个类目的类名，所以，杜威系统里的“焊接”类目是用类号“671.52”代表的。

枚举分类法没有标引词的合成措施，使我们只能用该系统中由先组法编定了的类号。所以，我们在标引时所能达到的专指程度就受到限制，我们不可能做到比分类法编者所编定的细目表更专指。通过先组，杜威在早期的版本里也在一定领域里达到了合理的专指程度。例如在第五版里(1894)，有一个类号“628.23”代表“下水道通风(ventilation of sewers)”。这个类号在第十六版里（1958）还照样保留着。类号“628.455”代表“城市垃圾处理(municipal garbage disposal)”，表示先组了的三个概念，是很专指的。枚举系统在个别等级(hierarchy)里可以做到专指，但由于缺乏词的合成措施，常常妨碍我们精确表达现代技术文献中所讨论的复杂的主题之间的关系。用《十进分类法》，我们可以用“671.52”表示“焊接(welding)”，也可以用“669.722”表示“铝（aluminum)”，但是我们无法确切表达“铝的焊接(welding of aluminum)”，因为杜威没给我们以将这两个类号合成的可能性。当然，我们可以把一本讲铝焊接的书，给出上述两个类号。但将这本书往书架上排的时候，就只能从这两个类号里选用一个。但是，在分类目录上，一条著录(an entry)可以在两个类目下都列上，以便我们无论循着“铝”这条途径还是循着“焊接”这条途径，都能查出这本书。这并不能让我们只检出讲铝的焊接的文献。为了确保在分类目录里就这个课题作到全面查找，我们必须查遍全部焊接的目录或全部铝的目录，如果我们对编目者是否总是作双重收录没有把握的话（图7）。因为这种系统是不能调配的（nonmanipulative），所以我们无法简便地选出恰好既属焊接类又属铝类的那组文献。目录上很可能有许多“焊接”类文献和许多“铝”类文献，而“铝的焊接”的文献却可能很少。查找一个准确课题（precise topic）时，由于为了检出少数有关文献的代号，而必须看许多无关文献的代号，所以是低效率的。

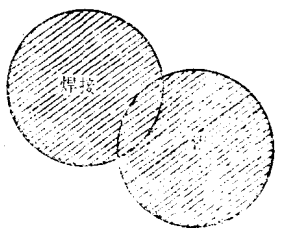

图7　“焊接”类同“铝”类的逻辑和

因为杜威也认识到了《十进分类法》基本上是枚举式的，试图为可能有的文献事先都给定一个现成类号是无益的事。所以他就准备了一定数量的合成措施。根据分类细目表里的各种细则，我们可以把一个类号合到另一个类号上，来表达这两个类号的逻辑积。这种合成通常叫做号码组配（number building）。地理复分可以看做是这种类型的合成措施。在分类法细目表的各部分里都有“复分如930-99”，它使我们可以用地理区域对一个课题复分。例如，550（地理学）可以进一步复分出“554.2”，即英格兰地理，英格兰在地理细目表上用942代

表*。《十进分类法》的其他一些地方都有“复分如表”的细则，这就使得我们得以将一个类号合到另一个类号上去，如：

016	目录
891.851	波兰诗
016.891.851	波兰诗目录

但在《十进分类法》里，合成成分受到严格限制。杜威警告说，除了有专门细则的地方以外，不要随便将类号结合起来。

《国会图书馆分类法》的类号大部分于1899到1920年间编出，几乎全部是枚举式的，实际上没有合成措施[7]。在该分类法里，虽然有些类（如，N类，美术）里包括有一个分国表可以附于细目表的一定类号上，但在许多类目里，即使是地理类目，也是分别枚举的。这个系统是高度先组的，许多领域里的专指程度达到了合理的水平。例如，在R类（医学）中，“RK510”代表“牙科麻醉”，而“RL 793”代表“皮肤先天异常”。这样，用国会图书馆的细目表就可能按类表现一篇文献的内容。但是这部分类法所达到的专指程度，也只能称之为“图书（专指）水平”，而不是，比方说，在科技期刊上可能遇到的文章的专指水平。

一部历史上有意义的，但在美国不大为人所知的图书分类法，是布朗，J. D. 的《主题分类法》。这部分类法于1906年首次发表（末版是第三版，1939 年出版）[3]。布朗的分类体制是高度合成式的。这代表着一种想要建立一种每一个主要论题只出现一次的“独位”(one place)分类法的努力。一个主要论题的各个方面可以用一个类目表(categoricae table)上的数码表示，该表列有“外形、状态、立场、合格性”等等，多多少少都能用到每一个主题上或主题的分支上。如，在代表“阿尔及利亚”的主类号“O 350”后，加上“移民”的类目号“.758”，就得出“阿尔及利亚移民”的代号“O 350.758”。同样可以把代表“炎症(inflammation)”的类号“.524”附加到“周围神经系（peripheral nervous system)”类号“G 703”后，来表示“周围神经炎症”。《主题分类法》中还没有类目的事物，布朗规定可以用主细目表里的两个标引词合起来，如：

F 237 E 026　　乌贼呼吸 (Cuttlefish respiration)

E 154 I 222　　修剪灌木 (Pruning of shrubs)

用得最多的基于合成原理的分类系统是《国际十进分类法(UDC)》[2]，是1905年在法国首次出现的。自那以后出版过各种语言的全文的和删节的版本。虽然《国际十进分类法》是在《十进分类法》基础上建立的，并且用的是同一大纲，但它却是高度合成式的。主要的合成手段是冒号组配，它可以用来把细目表里有的任何两个词结合起来（遇有为此目的专门设置的并优于冒号的手段时除外）。在《国际十进分类法》里，可以往“煤矿”的号码“622.33”上加上“通风”的号码“622.41”，而创立一个代表“煤矿通风”的新号码“622.33：622.41”。它有一种叫做“辅助细目表”，作为进一步复分的手段：

1. 形式：621.38(03)代表电子百科全书。
2. 地点：“德国青年运动”可用369.4(430)代表。
3. 语言：61(032)＝82表示俄文医学辞典。
4. 时间：327“1971”(42：43) 表示1971年英国和德国之间的国际关系。

* 该分类法第十七版里，这种类型的地理复分编成了一个“地区表(Area Table)”，其中以42代表不列颠群岛。

5．种族：378(＝924) 代表犹太人的高等教育。

《国际十进分类法》还在该系统的不同部分里采用了一种专用复分的辅助手段（包括连符），作为合成的一种手段。在农业部分里可看到一个例子，它的收成细目表633到635可以用连符连接农业问题细目表631到632而加以限定。如用连符将 633.11 “小麦” 同 632.7 “虫害” 连接起来，构成一个新的类号 “633.11-7”，代表 “害虫对小麦的为害”。密尔有两篇文章[8,10]对作为检索语言的《国际十进分类法》有杰出的分析，请参看。密尔早先著的一本教科书[9]，对于本章讲的各种分类系统都有充分详尽的讲解。

阮冈纳赞把分类法里的合成推进到符合逻辑的结论。1933年他首先发表了他的《冒号分类法》[11,12]。阮冈纳赞的系统建立在最低限度的枚举，并可随意将标引单元合成的真正自由的基础之上。阮冈纳赞的系统是一个分析-合成分类法，它的作法是：仔细地将每一主题领域分解成它的组面 (component facets)，把这些组面在细目表里排成有用的次序，并备有将这些组面结合（合成）来表示主题的任何程度的复杂性的规则。看一看《冒号分类法》的 L 类，医学：类号 “L 185” 代表 “眼”（来自解剖学的组面），而 “L 18517” 代表 “视网膜”，这是 “眼” 的分支。我们可以把代表 “炎症”（在疾病的组面里）号码 “415” 附加上去，就得出一个代表 “视网膜炎” 的号码 “L 18517：415”。

《冒号分类法》始终没得到多少应用，就是在它的发源地印度也如此。但阮冈纳赞的思想却在其他国家中广泛流传，特别是他的原理为伦敦的分类法研究小组的成员们所接受，编出了若干供专科领域（如教育、工程、制糖、包装、建筑业）使用的优秀的分类系统。

明显的是，分析-合成分类法如果编制得当，会提供实际上无限详细地表述文献内容的能力。图 8 是假想的设计技术领域里用的分类系统的一些方面的示例。可以把各个方面结合起来表示高度复杂的内容：如 “KpBcfi” 代表 “铬镍钢疲劳”，“LmKpBcf” 代表 “铬钢受冲击负荷的疲劳”，“KgbBqtAc” 代表 “镍合金管材的蠕变破裂”。

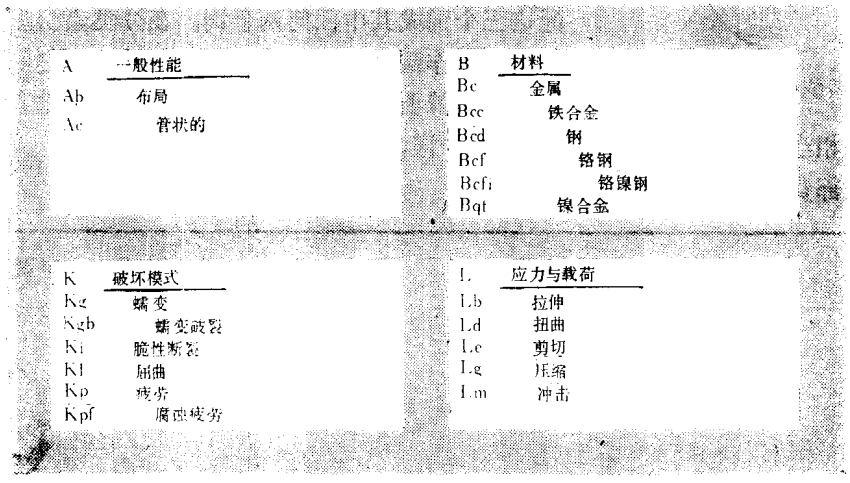

图 8　假想的设计技术领域里的分析-合成分类法系统的部分

分类系统通过合成办法可以提供实际上无限制的专指程度，因而能够处理极端复杂的内容。但是这些组织方案都是为在先组检索系统（通常是卡片式目录或印刷的书本式索引）中使用的。这类系统有一个严重的局限性：它们都是一维的，所以收录的款目必须编排成线性形式。

比如有一篇文献假定给的类号是 “BnDafiLx”，代表一个精确课题 “肺癌的放射治疗”，

其中“Bn”代表“肺”，“Dafi”代表“癌”，“Lx”代表“放射治疗”。因为这个主题款目是在分类目录里的，所以必定按标志中第一个单元“Bn”归档。这就使这份文献能与查找同“肺”有关的文献时被检出来，因为查找者查阅目录时要查“Bn”这个输入点(entry point)，或查与“肺癌”(“Bn”下再复分出“Dafi”)有关的文献。但是这条款目不能帮助查找者看到与“癌”有关的全部参考文献，因为“Dafi”在这里不是一个输入点(entry point)。我们可以想象到，这个索引里含有许多与“癌”有关的款目，但按照所损害的器官分散于全索引各处。这样，要查找全部与“癌”有关的文献就有点困难了。由于所采用的这种归档次序，要查找所有论述“放射治疗”文献的查找者也同样感到不便。

在先组式索引里，一个特定款目必须按其标志的第一个单元的次序归档，只能用重复款目的办法做到多途径检索。一个办法就是单纯地将一个款目轮排，使它的每个单元都成为输入点：

BnDafiLx

LxBnDafi

DafiLxBn

但轮排为人们提供的组配不一定都是有用的词。在前面引用的例子里，一个查找者可能检索“Lx”一词，查阅与“恶性肿瘤的放射治疗”有关的一切参考资料，而不管它的所在部位。因为“Lx”是按直接所损害的器官而不是按病理情况复分的，所以归档次序是起不到作用的。

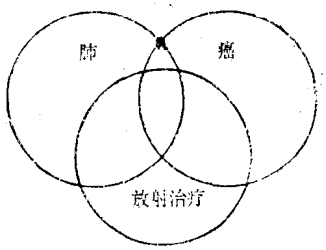

图9　后组系统词组配的各种可能

为了进行有效的情报检索，需要有对课题(topics)进行任意组配的灵活性。这点，在后组系统里能做到。后组系统里“癌”、“肺”和“放射治疗”都是独立的词，能用来进行各种组配（图9）。这样我们就有同等的便利去组配这三个词或其中任何两个词。在分类索引里，要使三个标志单元作到全部可能有的组配，就要把它们轮排，结果在档案里就要有六条款目。明显的是，轮排大大增加了索引的体积和相当多的费用*。即使款目轮排严谨，也会使体积加大，费用增高。下表列示的是轮排的全部费用（以档案厚度计）。

一份文献选用的标引词数	完全轮排的数
2	2
3	6
4	24
5	120
6	720
7	5,040
8	40,320

分类索引里为了避免款目的重复，而又仍旧保留多途径检索，阮冈纳赞[143]发展了链式

* Sharp[142]在SLIC索引里证明，标引词全部轮排可以避免，如果遵循确定的次序，数量较少的组配数就可以包罗一切途径。

标引法 (chain indexing) 的原理。用链式标引法，我们可以在分类档里只收单一标引词，而另编该档的字顺索引来提供另一种途径的标题索引。链式标引法是有系统地把指向所讨论的特定主题的概念等级链 (hierarchical chain) 的每级都标引出来*。用前面谈过的例子：我们是把那个款目归到分类档里 "BnDafiLx" 下的，它可以分解如下：

B	呼吸系统
Bn	肺
Da	肿瘤
Daf	肿瘤，恶性的
Dafi	癌
L	治疗
Lx	放射治疗

在链式标引过程中，我们自右至左把等级链中的每一级都标引出来，包含每一项有用途径的款目，按分类档的字顺排列如下：

放射治疗：癌：肺	Bn Dafi Lx
治疗：癌：肺	Bn Dafi L
癌：肺	Bn Dafi
恶性肿瘤：肺	Bn Daf
肿瘤：肺	Bn Da
肺	Bn
呼吸系统	B

链式标引不指引到具体文献，它引导用户去查分类档中能找到论及特定主题参考书目的那个部分。因此，链式标引中的款目可以指点出整整一组文献。每逢一份论述这个题目的文献加入这个系统的时候，我们不必次次都重复"癌：肺"这个索引款目。链式标引既不以惊人的速度增长，又能给我们以作到多途径检出主题的经济办法。

但是，即使在编制良好的链式标引里，尤其是如果分类档是特定主题领域里的高度详细的索引的话，会产生主题分散。链式标引法的这些局限性在克莱弗登写的第一批克兰弗尔德研究报告[4]中有充分的说明。

如果分类目录是以编制良好的、采用合成原理的分类体系 (classification scheme) 为基础，它就能提供高度专指的按学科检索的途径。但是因为这种分类档是先组的、一维的，所以它不能象后组索引法那样，给我们以自由组配一个课题的各个方面的灵活性。为了弥补这种缺陷，我们必须给复杂课题提供多种检索途径，或在分类档里重复款目，或给分类档加上字顺索引。

当然，为什么分类体系必须用先组法是说不出理由的。维克里[15]和兰开斯特[6]都提出来过，组配分类法非常适用于后组系统，罗塞[13]于1958年叙述过一种航空学科的以分类体系为基础的后组式系统。这里，各个标志单元分别选派给一份文献，使它们作为集体能代表文献的内容。查找时按照所想要的种种组配法用通常后组式的对照原理，例如计算机或光重合原理，把它们集合起来（见图10）。

* 链式标引法先决条件是分类号要严格坚持一个妥善规定的组面顺序，"著录次序"或"优先次序"来编定。

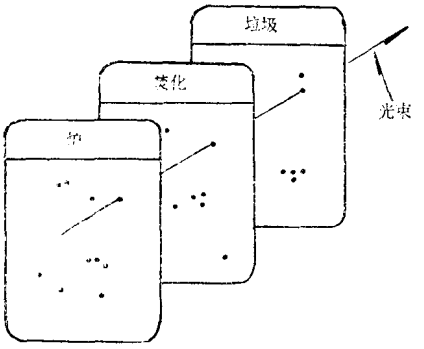

图1[1] 用光重合原理对照标志单元

分类体系当然能够用来进行同义词和同形异义词语义问题的规范化，并有族性检索的便利条件。试看英国电公司的《工程组面主题分类法》[1]的一角：

Qr	流体力学性质
Qrc	粘性
Qrd	可压缩性
Qre	弹性
Qrk	毛细作用，表面张力
Qrm	薄膜
Qrp	泡
Qrq	点滴生成，雾化

请注意，“毛细作用”一词同“表面张力”一词看作是十分密切的，作同义词处理，“点滴生成”同“雾化”也如此。这种近同义词的规范化作法是对词给出相同类号。这个分类法的字顺索引中，两个词都给同一类号。这就有效地防止了标引者把“毛细作用”同“表面张力”分开，并不管查找者脑子里哪个词先蹦出来，都能把论述该课题的文献检出来。

分类体系的编辑把一群代表流体力学性质的词集合到“Qr”组面里。这种集合有利于对流体力学性质进行族性检索。查找者无须想遍所有可能有用的词，因为在分类索引里，所有论及流体力学性质的文献标引词都有形地汇集在一起了。细目表各部分之间的互相参照项目也可用来便利族性检索。例如，在英国电公司的分类体系里，用参照项把“Kg 密封垫和密封盖”和它的相关词“油封和液封用 Kdad”联系起来。

同形异义词按它们在细目表中所处的上下关系以及它们在细目表中字顺的处理来区分，如：

Blasting: Cleaning Vvd

Blasting: Demolion Umze

在结束讨论分类体系之前，最后一个重要点是，分类法的标志只不过是类名的速记符号。除了保持细目表的秩序的功能之外，没有什么意义。从标引和查找的观点来看，把论及“焊接”的文献用一串英文字母“welding”标上或其他随便什么码号，如用“Vs”标上，都不重要。重要的是我们把什么作为一个类目（就是怎样限定它的范围），而不是把它叫做什么。

参 考 资 料

[1] Binns. J. and Bagley. D. A Faceted Subject Classification for Engineering. 3rd ed. Whetstone. England: English Electric Co., 1961.

[2] British Standards Institution. Guide to the Universal Decimal Classification (UDC), London, 1963. B. S. 1000 C: 1963.

[3] Brown. J. D. Subject Classification for the Arrangement of Libraries and the Organization of Information. 3d ed. London: Grafton & Co., 1639.

[4] Cleverdon. C. W. Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems. Cranfield, England: College of Acronautics. ASLIB Cranfield Research Project, 1962.

[5] Dewey. M. Dewey Decimal Classification and Relative Index. 17th ed. Lake Placid Club. New