

情报检索语言中的分面引用次序的研究

叶 千 军

一、分面公式的适用场合与意义

当对一个复杂主题划分了若干分面以后, 下一步就必须确定分面的引用次序, 也就是产生分面公式。这一步工作非常重要, 因为分面公式不仅仅适用于分面分类法, 而且适用于以下多种场合:

- 1) 体系法类目划分的次序
- 2) 叙词做先组使用时的次序
- 3) 轮排索引的编制

可以说, 引用次序是先组式语言(包括后组作先组使用)最为重要的特征。一个明确实用的引用次序体现着情报检索语言的句法规则, 它能够制约自然语言与情报检索语言之间词序的差异, 并进而提供:

A、复杂主题各因素之间明确的位置和含义

B、标引的一致性公式(必须认识一个主题有几个因素, 其组配次序就有 $n!$)

C、完整的检索策略公式(用户提问式往往很难完全)

所以, 引用次序对提高检索效率是至关重要的。这是因为:

- 1) 不同的引用次序决定着不同的体系结构, 它将直接影响检索效率。
- 2) 符合逻辑的引用次序将减少检索噪音(这可以从单元词无词序导致检索噪声的

缺点证实)

3) 引用次序一致将保证文献前处理的质量

因此, 引用次序的问题一直是情报检索语言研究的重点之一, 它又是分面分析的关键。侯汉清同志曾就阮岗纳赞、维克利、柯茨等人的研究作过归纳分析。^①本节不再作类似评述, 但拟从引用次序诸原则及其实施的问题作深入探讨。

二、引用次序的原则及其关系

凯塞、阮岗纳赞、柯茨、维克利的引用次序, 实际上都是按特定原则来排列实施的。如凯塞是按照“实体——过程”原则; 阮岗纳赞按照“具体性递减”原则; 柯茨是按照“重要性次序”原则; 维克利是按照“目的性”原则。除此以外还应包括最早注意引用次序的克特所提出的“通用术语”原则(或称为“自然语言”原则)。不同时期的原则留下了人们探索的足迹, 它是研究一步一步深化的标志。所以, 上述诸原则它们并不是互相对立, 而是互相联系, 互相制约的。这些原则从不同方面抓住了引用次序若干关键性问题。由此可以归纳出引用次序的原则为:

(一) 逻辑性原则; (二) 符合检索需要原则; (三) 表达性原则。

这些原则又表现为对引用次序技术上的

2. 《中国图书馆图书分类法》使用说明书 书目文献出版社 1981年11月

3. 《北京图书馆中文图书主题标引工作条例》

中文统一编目组 1984年10月

4. 《图书分类法》书目文献出版社 1983年9月

要求。即一个好的引用次序必须是既具有检索实用意义,又符合逻辑,易于表达和理解的。反之,则不能满足这些要求。但仔细观察就会发现,要同时满足这三项原则并不容易,因为它们之间可能是相悖的。如表达性强的词序可能与自然语言词序较一致,而与逻辑词序相违。因此必须明确这三项原则之间的关系是:以符合检索需要为前提,以逻辑次序为基础,而兼顾表达性原则。也就是说在逻辑次序与表达性之间,要以前者为主,并符合检索需要。下面分别研究三条原则的具体实施:

(一) 逻辑性原则

任何一个正确的分面公式都体现了一种逻辑次序。那么,逻辑次序究竟是什么样的次序?它通常有多少种?对于这问题目前尚未见文献专门论及。从表现上看,逻辑次序就是一种将各语义分面层层隶属,互相依附,各邻近分面之间具有紧密联系的顺序。换句话说,逻辑次序须具备两个要素:一是按语义因子分面;二是按具有相同义位词的集合构成关系语义场。其公式为:

$$\text{逻辑语义关系结构} = \text{语义因子} \\ + \text{逻辑句法结构}$$

从图一所列的二十种主题类型的引用次序,可以进一步看清逻辑关系的具体体现和范围。

图一基本上是按照柯茨提出的引用次序^②修改而成的。INIS(国际核情报系统)就是沿用这套逻辑次序制定的分面标引公式,使得分散在七十多个国家的标引仍能达到较高的一致性,检索效率比较好。我们由此可以进一步总结出如下规律:即逻辑次序乃是一种按照事物的发展顺序,符合基本思维规律,具有检索者心理依据和检索实用意义的次序。它与自然语言的表层结构不完全相同,但在大多数情况下,与人的本能是一致的,即人们通常是先提出一种事物,而后想到与

其联系的活动。先有目的、产品、研究对象,才有手段、工艺和研究方法。如果脱离了事物或环境,是很难想象它的各种活动的。这正如阮岗纳赞形象比喻的“壁画原则”,没有墙壁,就无法在壁上作画。但是,必须指出有时为了需要也会反过来作研究,即通过区分某种作用、功能、材料来研究某一事物或部件。如图二中的8—20就属于这类,而1—7则是从事物入手研究与之联系的性质、作用、材料或着眼点。

在以上各类逻辑关系之间还可进行综合或扩展。如:

A. 事物/材料 整流器/晒

扩为:

B. 事物/部件/材料/性质

整流器/晒/铸件/规格

同样,一个详尽的逻辑组面也可分析成若干对简单的逻辑次序,甚至仅抓住关键的关系,省略次要的。这个问题与各分面表达主题的重要程度与联接能力有关。实践证明,情报检索语言所采用的引用次序是一个“具体性递增,必要性递减”的次序。^③较后层次的主题因素相对于较前层次的主题因素来说,在描述文献主题上的必要性较小。而这种必要性又是由各个面的联接能力决定的。阮岗纳赞认为,各分面按联接能力上升排列其序列为:空间、时间、动力、物质、本体。^④这些重要理论,为情报检索语言编制和使用中专指度的调节提供了理论依据。它具体适用于:

- 1) 机检或手检不同专指度的调节;
- 2) 分类表或词表划分层次的选择;
- 3) 书本式索引的编制。

应该指出:无论怎样分析和综合,不管面的数量多少,在组成一个标识串时,都必须贯彻逻辑原则。衡量一个较为复杂的引用次序是否符合逻辑次序有两种判别方法:

第一是“截除法”。即通过截除一个主题中较次要的主题因素来判断。截除的方法

表一 逻辑关系和引用公式*

	逻辑关系类型	引用公式	主 题	标 引
1	事物的作用	事物——作用	隧道开挖	隧道/开挖
2	材料的作用	材料——作用	铝的铸造	铝/铸造
3	A作用施加于B作用	B作用—A作用	辐射对新陈代谢的影响	代谢/辐射
4	事物的材料	材料——事物	磁芯体的钢材	钢/磁芯体
5	事物的部分	事物——部分	自行车轮胎	自行车/轮胎
6	事物、材料、作用的特性	事物 材料 作用 } 特性	A. 鼠类的繁殖力 B. 混凝土的强度 C. 腐蚀的速度	A. 老鼠/繁殖力 B. 混凝土/强度 C. 腐蚀/速度
7	事物、材料、作用或特性的某个着眼点	事物 材料 作用 特性 } 着眼点	A. 加速器的安全 B. 铀矿物化学 C. 宣传的费用 D. 寄生振荡的频率	A. 加速器/安全 B. 铀矿物/化学 C. 宣传工作/费用 D. 寄生振荡/频率(物理)
8	按作用原理区分事物	事物——作用	悬吊式桥梁	桥梁/悬吊式
9	按材料、能源形式区分事物	材料 事物 能源形式	A. 硒整流器 B. 电动马达	A. 整流器/硒 B. 发动机/电
10	按部件区分事物	事物——部件	晶体管脉冲积分器	脉冲积分器/晶体管
11	用B作用区分A作用	作用A—作用B	材料的爆炸成型	材料加工/爆炸成型
12	A事物向B事物提供服务或以B事物为目的	A事物—B事物	A. 电视机用稳压器 B. 回旋加速器用氚核源	A. 稳压器/电视机 B. 氚核源/回旋加速器
13	事物对作用的服务或手段	作用——事物	植物生长的用水	植物生长/水
14	B事物引起A事物	B事物—A事物	石坠落的坑	石/石坑
15	作用引起事物	事物——作用	用冷凝法生产晶体	晶体/蒸汽凝聚
16	事物引起作用	事物——作用	海水引起解吸	海水/解吸
17	B作用引起A作用	A作用—B作用	电解引起的腐蚀	腐蚀/电解
18	特定场合的某事物	场合——事物	沙漠中发现的植物	沙漠/植物
	特定场合的某作用	场合——作用		
19	事物与地区或机关团体	事物——地区	中国的人口控制	人口控制/中国
20	事物与文献类型	事物——文献类型	气象预测会议录	气象预测/会议录

* 此表是按E. J Coates的引用公式修改而成——作者注

表二 相关关系及其符号

关系实质	基本大类之间	分面之间	同位类之间
一般关系	a	j	t
倾向关系	b	k	u
比较关系	c	m	v
差异关系	d	n	w
工具关系	e	p	x
影响关系	g	r	y

是从后往前,依次省略处于末后层次的主题因素。如果截除后的结果只是标识的专指度降低,主题的基本方向并不变,则可判断为是正确的应用次序。反之,则是错误的。

如主题:春季奶牛的肠渗透压

标引: A 奶牛/肠/渗透压/春季

B 渗透压/春季/肠/奶牛

截除: A₁ 奶牛/肠/渗透压

A₂ 奶牛 / 肠

B₁ 渗透压 / 春季 / 肠

B₂ 渗透压 / 春季

分析: A为正确的引用次序, B为错误的。因为A通过“截除法”检验, 当依次省略处于末后的第四, 第三个主题因素, 分别用A₁和A₂标引时, 主题的原范围不变, 只是专指度有所降低; 而B则正好相反, 照此顺序省略后, B₁和B₂均显得模糊不清, 无法表达文献的主题内容。

第二是“链环检查法”。正确的引用次序犹如一串有序的链环, 一环扣一环, 不容脱漏, 错环, 也就是必须循序渐进地依次划分。如果采取跳跃式引用, 如下例就是错误的。

物质 / 成份 / 部份 / 性质

因为它破坏了一个本应紧联着的微观逻辑关系。用阮岗纳赞的比喻“牝牛——牛犊”原则, 即奶牛及其未断奶的牛不应分开。B例“奶牛”和“肠”分开就属这类错误。

(二) 检索需要原则

不同学科主题面的数量和引用次序不同, 而引用次序所贯彻的检索需要原则, 其实质就是提供在通用次序公式指导下的次序变换问题。

一个引用次序的确定与主题中心的确认有直接关系, 而主题中心通常是因学科或特定的检索对象而异的, 也就是说特定的检索者有其相应的检索重点和方式。如“金霉素治疗眼结膜炎”这个复杂主题, 眼科医师提出的逻辑次序可能是:

眼: 结膜炎: 金霉素

(器官) (疾病) (药物)

而药剂师的提问则可能是:

金霉素: 结膜炎: 眼

(药物) (疾病) (器官)

这两种引用次序不能说哪一种更好, 只有分别满足不同的检索需要才是正确的。而这个问题的关键之一又是首先要确定主题中心。主题中心是最关键的主题要素。它相当

于本体面, 直接承担检索功能。在逻辑语义的主题因素中可称作“逻辑语义核”。而没有承担检索功能的其它主题因素可称作“非逻辑语义核”。它们的位置往往是根据与主题中心的逻辑关系及其亲疏来确定的。物质、动力、空间、时间被作为一种具体的研究事物时, 都可能成为本体——主题中心。这就是不同的检索需要会有不同的主题中心。而依据主题中心确定的相应次序又是符合特定专业大多数读者需要的次序。从表一也可看到, 这二十种逻辑关系次序基本上是从确认一个主题中心的前提下进行的, 其中很多类型都可交替变换主题中心。衡量一个标引工作者水平方面之一就是看他是否能够反复自问, 这篇文献会有哪些合理的提问, 然后通过表一这类公式查明其逻辑关系次序。

除了主题中心后移的现象对于确认主题中心有作用外, 其次运用语言学中的“中心成分分析法”“语法分析法”(即分主、谓、宾结构)也有一定作用。但这些方法有一定弊病, 即并不考虑特定检索对象不同的主题中心, 所以这里即不作专门讨论。但鉴于主题中心对于检索和引用次序的特殊重要性, 再研究一下究竟哪一些词可作为检索入口词以及确定它的原则也是必要的。在这方面应该注意三点:

1) 优先选用与本检索系统直接相关的专业主题作款目词。

2) 款目词应是比较专指, 且具有检索意义的。下列概念不应作为款目词:

A、过于宽泛的概念。如: 教育——农村 历史——泰国

B、仅具有构词功能的概念。如: 袖珍式、多功能

3) 款目词应代表文献研究的对象。各种文献主题以外的外部特征, 如文献形式、著者、出版单位等, 当不作为研究对象或起类集文献作用时, 均不应该作主题中心处理, 以免引起检索噪音。如下例:

A. 《关于文摘的编制》 标为:

文摘/编制 (正确)

B. 《图书馆学文摘》 标为:

文摘/图书馆学 (错误)

当从一串较复杂的主题因素中,明确一个新的主题中心之后,次序的重新排列问题也应有一条原则,即:一个复杂主题不论被不同的检索对象确认为几个主题中心,在具体的一个引用次序中,都只有一种正确的表达方法。任何引用次序都必须以逻辑次序为基础。标记符号的使用对于改变引用次序具有明显作用。张琪玉总结的三种方法^⑤具有实用意义。这就是:

I、DGJLA

A、分两段法:ADGJL改为: II、GJLAD

III、LADGJ

B、逐个倒置法:ADGJL改为: I、D、A-GJL

II、G、D、A-JL

C、分若干段法:ADGJL改为: I、G、AD-JL

II、L-J-ADG

以上方法均可起到改变主题中心,增加聚类方向和检索入口的作用。除此以外任意的调换位置都被认为是不可能的。

(三) 表达性原则

一般来说,控制语言的表达性总是不及非控制语言的。因为非控制语言可以通过:

1) 介词、虚词、形容词等一系列语助词,
2) 与人们理解一致的词序,这两个有利条件达到较好的表达性。而控制语言则正好相反: 1) 对词类进行选择,一般只收名词和名词性词组, 2) 与自然语言词序不完全一致。由于这些差别存在,情报检索语言中由于错误组配造成的“主题失真”现象是经常发生的。它对于检索效率的影响是严重的。这种现象,即存在于语词标识,当后组语言作先组使用时,也反映在号码标识使用时。但以语词标识的问题显得更重要,更迫切。因为在我国后组语言作为过渡正推荐作先组使用。提高组配标题的易懂性,是当务之急。

解决这个棘手问题的主要手段有两条: 1) 通过一致的引用次序, 2) 运用有效的标识符号,来清楚地区分各种组配的类型,从而提高控制语言的表达性。

在解决这个问题之前,充分地揭示出哪些组配通常易造成主题失真十分必要的。据目前应用分析,以下组配次序应引起注意。

A、某些材料与产品 如: 用球墨铸铁制的齿轮

标为: 轮——球墨铸铁(也会被理解为: 齿轮的球墨铸铁,即表达为事物与方面之间的关系)

B、某些设备与用途 如: 加油飞机

标为: 飞机——加油(会被理解为事物与行为、操作)

C、某些整体与部分 如: 润滑系统中所使用的泵

标为: 泵——润滑系统(会被理解为“泵的润滑系统”)

D、影响与被影响关系 如: 鲁迅对郭沫若的影响

标为: 郭沫若——鲁迅——影响(难以分清影响与被影响)

但是又必须指出,这类表达不准确的组配有时对检索效率的影响并不是很大。这是因为一种两可的组配含义,当遇上懂专业的检索者则不一定会造成歧义,他们能够判断出一个概念的正确理解。二是借助一致的引用次序以及职能,关系符号等,可大大防止或减弱检索“噪音”。

参考文献

① 侯汉清《试论主题标引中的引用次序》载《情报学报》1985年第3期

② 柯茨著《主题目录: 标引与结构》1960年

③⑤ 张琪玉《文献主题的构成因素及层次》载《图书情报知识》1985年第1期

④ (印)塞克德瓦著,吴人珊译《冒号分类法理论和实践》书目文献出版社1985年