

情报检索语言

武汉大学图书馆学系编

1980 . 6 .

80-814

THE
LIBRARY
OF THE
MUSEUM OF
ARTS
AND
ARCHAEOLOGY
OF THE
UNIVERSITY OF
TORONTO

1975

1975

1975

书 号 G354.41/WDT

登记号 220215

从借出日
期起限借
阅一个月

220215



前 言

《情报检索语言》这门课程是根据我系 1978 年 11 月制订的科技情报专业和图书馆学专业四年制学分制教学方案开设的。

情报检索是情报传递中的重要环节。加强这部分知识的讲授，一方面要增设有关建立电子计算机检索系统的一系列课程，另一方面还要增设有关情报检索语言的课程。当前，作为一个科技情报专业或图书馆学专业的毕业生，仅仅掌握体系分类法一种类型的情报检索语言已显得不够，而需要掌握更多类型的情报检索语言知识，以便走上工作岗位后，对各种语言能正确、熟练地运用，在实践中不断加以改进，并结合具体需要创制新的情报检索语言。

各种类型的情报检索语言虽有不少差别，但在基本原理上是一致的。当前情报检索语言的发展趋势是互相渗透，各种方法互相采用。因此，将各种类型的语言合在一门课程中，以提高其检索效率为重点，对其构成原理、方法和性能等作系统的对比式讲述，可能是有益的。这门课程就是在此想法指导下的一个尝试。

这门课程系初次开设，应包括哪些内容，怎样进行讲授，这些我们都还刚开始探索。在编写讲义时，参考资料亦感不足。加之，水平有限，所以，这本讲义可以说只是初具雏形，缺点错误一定较多，印出试用，并请同志们多加指正。

这本讲义的编写者是张琪玉同志。

武汉大学图书馆学系

1980年3月17日



A0073015

I

W D T

目 录

第一章 绪论	1
第一节 情报检索语言在情报检索中的作用	1
第二节 对情报检索语言的要求	2
第三节 概念逻辑和科学分类是情报检索语言的基础	6
第四节 情报检索语言的分类	11
第二章 等级体系分类语言	13
第一节 体系分类法的原理	13
一 体系分类法的构成原理	13
二 类目的划分	15
三 类目的排列	19
四 类目的名称及涵义	24
五 学科之间相互关系的处理	27
六 某些图书资料分类的特殊性	30
第二节 体系分类表	31
一 分类表的功用和结构体系	31
二 分类表的编制法	33
三 分类表的设计和编程序	50
四 分类表的使用法	51
五 分类表的管理	53
第三节 体系分类法的性能	55
一 体系分类法的检索效率	55

二 体系分类法的某些问题.....	56
第三章 分析—综合分类语言.....	60
第一节 组配分类法的原理.....	60
一 组配分类法的构成原理.....	60
二 “面”的划分和细分.....	64
三 类号的组配和排列.....	66
四 组配分类法的类型.....	67
第二节 组配分类表.....	68
一 组配分类表的结构体系.....	68
二 组配分类表的编制法.....	69
三 组配分类表的使用法.....	71
第三节 组配分类法的性能.....	72
一 组配分类法与体系分类法的比较.....	72
二 组配分类的某些问题.....	74
第四章 标题词标引语言.....	76
第一节 标题法的原理.....	76
一 标题法的构成原理.....	76
二 标题的几种形式.....	77
三 标题词的范围及选词原则.....	79
四 标题词的规范.....	81
五 标题词相互关系的显示.....	85
第二节 标题表.....	90
一 标题表的功用和结构体系.....	90
二 标题表的编制法.....	91
三 标题表的使用法.....	97

四	标题表的管理	99
第三节	标题法的性能	100
一	标题法与体系分类法的比较	100
二	标题法的某些问题	102
第五章	单元词描述语言	104
第一节	单元词法的原理	104
一	单元词法的构成原理	104
二	单元词的范围及选词原则	105
三	单元词的规范	106
第二节	单元词表	109
一	单元词表的功用、结构体系和编制法	109
二	单元词表的使用法	110
第三节	单元词法的性能	111
一	单元词法与标题法、组配分类法的比较	111
二	单元词法的某些问题	113
第六章	叙词描述语言	116
第一节	叙词法的原理	116
一	叙词法的构成原理	116
二	叙词的范围及选词原则	120
三	叙词的组配	123
四	叙词的规范	133
五	叙词相互关系的显示	142
第二节	叙词表	151
一	叙词表的功用和结构体系	151
二	主表的编制法	154

三	辅表的编制法.....	16
四	叙词表的设计和编程序.....	16
五	叙词表的使用法.....	17
六	叙词表的管理.....	17
第三节	叙词法的性能.....	17
第七章	关键词描述语言.....	17
第一节	关键词法的原理.....	17
一	关键词法的构成原理.....	17
二	关键词法的形式.....	17
三	关键词的范围.....	18
四	关键词的规范.....	18
第二节	关键词法的性能.....	18
第八章	文献分析与标引问题.....	18
第一节	文献的分析与标引是影响情报处理质量的关键 性工序.....	18
一	保证分析标引质量的重要性.....	18
二	评价分析标引质量的标准.....	18
第二节	文献分析与标引的一般方法和规则.....	18
一	分析水平与标引深度.....	18
二	分析角度.....	18
三	分析方法.....	18
四	标引方式方法.....	18
五	导致检索效率降低的分析标引误差.....	19
第三节	文献分析与标引的某些问题.....	19
一	适度标引问题.....	19
二	集中标引问题.....	19
三	利用已有标引成果问题.....	19
	参考文献目录.....	19

第一章 绪 论

第一节 情报检索语言在情报检索中的作用

情报检索语言是根据情报检索的需要而创制的人工语言，简称情报语言或检索语言。

情报检索是情报传递中的重要环节。为了使情报使用者获得所需的情报，就要保证情报处理人员对情报内容的表达（标引用语）和情报使用者对相同内容的情报需求的表达（检索用语）取得一致，即他们必须使用一种共同的语言，而且这种语言必须是很严密的，才能顺利实现对情报的检索。情报检索语言就是这样的语言。

情报检索的全过程包括情报的存贮和检索两个方面。情报存贮是指编制检索工具和建立检索系统；情报检索即是利用这些检索工具和检索系统来查寻所需的情报。所以，情报的存贮和检索是二个有着密切联系的过程。当存贮情报时，情报处理人员首先要对作为情报载体的各种文献进行主题分析，即把它所包含的情报内容分析出来，使之形成若干能代表文献主题内容的概念，并用情报检索语言的“语词”（标识）把这些概念标示出来，然后纳入检索工具或检索系统。当检索情报时，情报使用者也首先要对自己的检索课题进行主题分析，形成若干能代表情报需要的概念，并转换成情报检索语言的“语词”，然后从检索工具或检索系统中查寻正是用该“语词”标引的文献，从而找到他所需要的情报。由此可见，情报检索语言是情报处理人员和

情报使用者所必须共同遵守的规范。

情报检索语言可以从自然语言中精选出来并加以规范化的一套词汇，用它来对概念进行描述；可以是代表某种分类体系的一套分类号码，用它来对概念进行逻辑分类；也可以是代表某一类事物的某一方面特征的一套代码（如化合物分子式代码），用它来对该类事物从某一方面进行标引。

情报检索语言在情报检索中的功能大致可归纳为三点。

1. 对文献的情报内容（及某些外表特征）起标识（揭示）作用；
2. 由于这些标识都有一个固定的排列次序，所以能将情报加以系统化（或组织化）；
3. 由于它的语词或符号具有单义性和相互联系，所以能将内容相同或相近的情报进行集中或揭示其相关性。

情报检索语言的质量以及如何正确使用，对情报检索效率起着决定作用。所以，如何提高情报检索的效率（特别是检全率和检准率），是现今情报检索语言研究的主要目标。

第二节 对情报检索语言的共同要求

检索效率是由许多因素构成的，所以对情报检索语言也有多方面的要求：

1. 要保证检全率和检准率高，或者说，漏检率和误检率低。检全率是指对各种文献用某种情报检索语言标引并纳入检索工具或检索系统之后，在需要时能将其中载有有关情报的那些文献全部检出。事实上，很难达到全部检出的要求，而只能达到某个百分比，

就会有一部分有关的文献未能被检出，这一部分文献的数量可用漏检率来表示。与某项情报需要有关并且被检出的文献量和虽有关但未能被检出的文献量之和，等于有关该项情报的全部文献量。由此可见，检全率和漏检率是可以说明同一问题的两个相对的概念。检全率高则漏检率就低，而漏检率高则检全率就低；如果检全率为60%，则漏检率就是40%。检全率是情报检索质量的主要标志。因为漏检率高，表明重要情报被遗漏的危险性大，所以在情报检索中，总是力求提高检全率，或者说降低漏检率。

在情报检索中，当检出与某项情报需要有关的文献的同时，常常会带出一些无关的文献。事实上，要完全避免带出无关文献也是很难达到的。所以检准率也必然是某个百分比。换句话说，会存在一定的误检率。与某项情报需要有关并且被检出的文献量和虽无关但也被带出的文献量之和，等于被检出的全部文献量。由此可见，检准率和误检率也是可以说明同一问题的两个相对的概念。检准率高则误检率就低，而误检率高则检准率就低。因为误检率高表明用于甄别所检出文献是否真正有用的无谓劳动的增加和时间的浪费，虽不会降低检索质量，但却会影响检索效率，所以在情报检索中，总是力求提高检准率，或者说降低误检率。

保证较高的检全率和检准率，是对情报检索语言的主要要求。

2. 易于标引，易于检索。情报检索语言语词或符号含义的明确性，整个语言包含概念的丰富和完备性，语言结构上的严密性，标识的直观性（语言符号及其排列次序易于理解），等等，所有这些，可使标引者易于掌握，标引速度快，标引分歧少；检索者也易于掌握，检索时间短，检索差错少。

3. 具有对多种检索要求的适应性。例如。族性检索、特性检索、

扩检、缩检、多途径检索、引进布尔代数表达式等。族性检索一是范围较广泛的检索，常常是指从学科专业出发的检索，或者是一系列同族事物的检索。例如，检索图书馆学范围的文献或检索于所有各种类型图书馆的读者服务工作的资料。特性检索一般是范围较狭窄的检索，常常是指从某种事物出发的检索，或者是对某特定事物甚至是对其某一特定方面、特定问题的检索。例如，检关于某种类型飞机的全部资料或某一型号飞机飞行性能的资料，等等。一般认为，分类语言比较适于从学科或专业出发进行的族性检索，描述语言比较适于从事物或问题出发进行的特性检索。实际在族性检索和特性检索之间不可能划一个明确的界限，任何情报检索语言都应采取多种措施，力求既适于族性检索的要求，也适于特性检索的要求，并具有随检索中出现的具体情况而自由或缩小检索范围的灵活性。多途径检索是指一个复合概念的资料从多条途径，手都可检出。例如，关于开合式铁路钢桥的资料，从铁路桥、开式桥或钢桥三条途径入手都可检出。布尔代数即逻辑代数，用以准确地表达检索课题和进行逻辑运算，提高检索质量。

4. 具有对先进的检索方式和检索设备的适应性。检索方式分文献单元方式和标识单元方式。检索设备如普通卡片式目录、本式目录、比号卡片、比孔卡片、穿孔卡片、机械式检索系统、电式检索系统、电子计算机检索系统等。各种情报检索语言对各种检索方式和各种检索设备的适应性各不相同。例如，体系分类法于电子计算机检索系统就不能充分发挥计算机功能，上下文关键法适用于电子计算机系统却不适用于卡片式目录，单元词法适用比号卡片和比孔卡片却不适用于普通卡片式目录，等等。当前，情报检索的机械化自动化是一个发展方向，所以，特别要考虑情报

索语言对电子计算机等先进检索设备的适应性。例如，叙词法和组配分类法就能充分发挥电子计算机的功能。

5. 具有对图书馆或情报机构各个环节的适应性。在这方面，体系分类法（包括体系组配分类法）的适应性最强，除适用于情报检索外，还适用于排架、藏书或服务部门的划分、文献报导、统计、藏书补充计划等。其它语言则主要适用于情报检索和文献报导。

6. 对文献学科和类型的适应性，以及对某种类型图书馆和情报机构的适应性。一般地说，情报检索语言的专业性或专用性强，则在处理本专业或特定类型的文献上效率比较高，但在处理其它专业或其它类型文献上效率比较低，甚至完全不适用。综合性的情报检索语言，适用于综合性图书馆和情报机构处理各种学科和类型的文献，但用于某一专业单位或某一特定类型文献时，就不如各种专业的或专用的情报检索语言效率高。这是一个矛盾。如何使它兼备两者的优点，尚是一个努力方向。

7. 与其它情报检索语言的兼容性以及一国通用性和国际通用性。每一种情报检索语言都是根据一定需要而编制的，不但具有适应某种具体需要的特点，而且往往还带有国家特点和自然语言特点。这样，造成各种检索工具和检索系统互不通用，不能互相利用标引成果，检索也很不便。所以，力求一国通用性和国际通用性。近十年来，由于电子计算机在图书情报工作中应用的进一步发展，已越来越多地建立电子计算机情报检索网络。由于联网的需要，更迫切要求解决各种情报检索语言的兼容性问题。各个检索系统都使用同一种情报检索语言是不可能的，因为至今还没有大概也不可能有一种情报检索语言能满足一切实际需要。但是，如果能使各种情报检索语言特别是同一类型的情报检索语言接近起来（例如尽可能用同一个语词来表达同一个概

念, 等等), 就有助于克服障碍和提高效率。

8. 具有不断“现代化”的可能性。即要能跟上科学技术和社会发展。科学技术领域中新科学新事物不断出现, 各科学之间的关系在不断变化, 社会也在不断发展, 情报检索语言要易于扩充修改及时增补新概念, 反映新事物, 以及改变旧的结构和概念联系, 不断进行“现代化”改造, 才能适应情报检索的需要, 否则将成为僵死的东西而被淘汰。

第三节 概念逻辑和科学分类是情报 检索语言的基础

任何一种情报检索语言, 不论是语词的还是符号的, 都是表达一系列概括文献情报内容的概念及其相互关系的标识体系。因此, 它们全都是建立在概念逻辑的基础上的。

概念逻辑是一种科学思维, 它能揭示事物的本质属性及各种事物之间的联系和区别, 概念则是事物本质属性的概括。所以, 文献的情报内容只有用概念才能加以科学的说明, 即通过主题分析将其概括为某个或某些概念。正象在日常的思想交流中是用各种自然语言来表达概念一样, 在情报检索过程中则是用各种情报检索语言来表达概念的。为了既能检索具有某种或某些属性的某一事物的情报, 又能检索具有同一属性或相近属性的一系列事物的情报, 情报检索语言不仅要表达一个个不同的概念, 而且还要显示各种概念之间的逻辑关系。

为了了解情报检索语言的基本原理, 我们需要具备一些概念逻辑的一般知识。

什么是概念？上面已经指出，它是事物本质属性的概括。毛泽东同志说：“概念这种东西已经不是事物的现象，不是事物的各个片面，不是它们的外部联系，而是抓着了事物的本质，事物的全体，事物的内部联系了”。例如，有一种机构，它从事图书资料的蒐集、整理、保存并以借阅方式提供利用。我们在对这种机构的本质属性进行概括之后，就形成一个概念，即“图书馆”。

概念有内涵和外延。概念的内涵是它所指事物的本质属性的总和，即概念的涵义。例如，“从事图书资料的蒐集、整理、保存并以借阅方式提供利用的机构”这些本质属性的总和就是“图书馆”这个概念的内涵。概念的外延是它所指的一切事物，即概念的适用范围。例如，“图书馆”这个概念适用于一切“从事图书资料的蒐集、整理、保存并以借阅方式提供利用的机构”。

概念的内涵有深浅，即它所概括的事物本质属性有多少；外延有广窄，即它的适用范围有大小。概念的内涵和外延之间成反变关系。概念的内涵越浅（即它所概括的事物本质属性越少），则它的外延越广（即它的适用范围越大）；反之，概念的内涵越深，则它的外延越窄。例如，在“图书馆”这个概念的内涵中增加“为科学研究服务的”这个属性，就形成“科学图书馆”这个概念，如果再增加“以社会科学的图书资料为范围的”，就形成“社会科学图书馆”这个概念。“科学图书馆”这个概念只能适用于一部分图书馆，而“社会科学图书馆”这个概念只能适用于一部分“科学图书馆”。

概念之间的关系，可以分为相容关系和不相容关系。如果两个概念的外延至少有一部分相重合，则两者之间是相容关系，互称为相容概念。如果两个概念的外延没有一部分相重合，则两者之间是不相容关系，互称为不相容概念。

在相容关系中又可以分为同一关系、属种关系、交错关系、整体与部分的关系和全面与某一方面的关系。在不相容关系中又可以分为并列关系、矛盾关系和反对关系。

同一关系，又称重合关系。是指有同一外延而又有不同内涵的两个概念之间的关系。例如，“北京”与“中国首都”这两个概念的关系即为同一关系，它们互称为同一概念。

属种关系。是指这样两个概念，其中一个概念完全被包括在另一个概念的外延里，是另一个概念外延的一部分。例如，“文学作品”与“小说”这两个概念的关系就是属种关系。就包括另一个概念的概念（外延广的概念）对被它包括的概念（外延窄的概念）来说是上位概念，或称属概念；就被包括的概念对包括它的概念来说是下位概念，或称种概念。例如，“文学作品”是“小说”的上位概念，即属概念；“小说”是“文学作品”的下位概念，即种概念。属种关系是概念从属关系的基本形式。

交错关系，又称交叉关系。是指有一部分外延重合的两个概念之间的关系。例如，“彩色电影”与“宽银幕电影”这两个概念的关系就是交错关系。即彩色电影中有宽银幕的，宽银幕电影中有彩色的。这样的两个概念互称为交错概念（或交叉概念）。两个概念外延的重合部分是一个新概念，其内涵等于两个概念内涵之和（即“宽银幕彩色电影”或“彩色宽银幕电影”），它对原来两个概念的任何一个来说，都是下位概念。

整体与部分的关系。是指这样两个概念，其中一个概念概括表示某一事物，而另一个概念则仅仅表示该事物的某一个部分。例如，“人体”与“人的心脏”这两个概念的关系就是整体与部分的关系。就“人体”对“人的心脏”来说是整体概念，就“人的心脏”对

“人体”来说是部分概念。整体与部分的关系是概念从属关系的另一种形式，整体概念是上位概念，部分概念是下位概念。但是，这种关系与属种关系不同。属种关系是就两个事物之间属性上的相同点及相异点而言；而整体与部分的关系则是就某一事物与其自身某个局部的关系而言。

全面与某一方面的关系。是指这样两个概念，其中一个概念概括表示某一事物的全部问题，而另一个概念则仅仅表示该事物某一个方面的问题（例如表示某一产品的原理、设计、制造材料、设备、工艺、使用方法等等之中某一方面的问题）。全面与某一方面的关系也是概念从属关系的一种形式，全面概念是上位概念，某一方面概念是下位概念。这种关系同整体与部分的关系很相似，而与属种关系显然不同。

并列关系，又称同位关系。是指同一个上位概念（属概念）之下的几个下位概念（种概念）之间的关系。例如，“省图书馆”、“市图书馆”、“县图书馆”这几个概念的关系就是并列关系，它们都是“公共图书馆”这个下位概念之下的上位概念。具有并列关系的各个概念，互称为并列概念或同位概念。在同一整体概念之下的各个部分概念之间，以及同一全面概念之下的各个方面概念之间，也都是并列关系。并列关系的各个概念在外延上是没有任何一部分重合的。

矛盾关系。这是并列关系的一种特殊形式，是指外延总和等于其上位概念全部外延的两个并列概念之间的关系。例如，“金属材料”和“非金属材料”这两个概念之间的关系就是矛盾关系，它们互称为矛盾概念；它们的外延之和等于“材料”这个上位概念的全部外延。

反对概念，是指两个互相并列而且互相对立的概念，但两者外延之和小于其上位概念的外延。例如，“导体”和“绝缘体”这两个概念之间的关系就是对立关系，它们互称为对立概念；它们的外延之

和小于“材料导电性”这个上位概念的外延，因为除此之外，还有“半导体”。

情报检索语言在表达各种概念及其相互关系时，普遍地应用了上述概念逻辑的原理，并且有效地利用了概念划分和概念组配这二种逻辑方法来建立自己的结构体系。

概念划分是利用概念内涵由代表事物属性的概念因子构成，概念因子的增加或减少可以形成新的概念，概念内涵与外延成反变关系等的性质，对概念进行划分或概括，形成更为专指或更为泛指的新概念，用以区别客观世界千差万别的事物；并利用划分结果所产生的概念从属关系和并列关系，建立某些形式的情报检索语言结构体系——概念等级体系，用以显示客观世界千差万别的事物之间的内在联系。体系分类法是应用概念划分这种逻辑方法的典型。

概念组配是利用在概念的交错关系中，两个概念外延重合处可以出现一个新概念，其内涵等于原来两个概念内涵之和，并且它与原来的两个概念具有从属关系的这种性质，进一步发展为将一个内涵较深（多）的概念分析为（分别归属于）二个或三个内涵较浅（少）的概念，以及将二个或三个内涵较浅的概念综合（归併、合成）为一个内涵较深的概念的一种概念逻辑方法，用以建立另一些形式的情报检索语言结构体系——概念组配体系。这种结构体系可以提供从多种途径来进行情报检索的条件，并且可以任意选择和改变标识的专指度，随具体情况扩大或缩小检索的范围。组配分类法和叙词法是应用概念组配这种逻辑方法的典型。

科学分类是对客观世界千差万别的事物之间本质联系和区别的更为深入的揭示，因而是显示各种概念之间相互关系的最系统的方法。毛泽东同志说：“科学研究的区分，就是根据科学对象所具有