



中央广播电视大学  
图书馆学专业用书

# 情报语言学基础

张琪玉 编著

TUSHUGUANXUE

武汉大学出版社

中央广播电视大学  
图书馆学专业用书

# 情报语言学基础

张琪玉 编著

武汉大学出版社

# 目 录

前言	( 1 )
第一版前言	( 1 )
第一章 绪论——情报语言学一般问题	( 1 )
第二章 情报检索语言基本理论	( 6 )
第一节 情报检索语言在情报检索中的作用	( 6 )
第二节 对情报检索语言的共同要求	( 9 )
第三节 概念逻辑和知识分类是情报检索语言 的基础	( 17 )
第四节 文献主题的构成因素及层次	( 29 )
第五节 情报检索语言的分类	( 34 )
第三章 等级体系型分类检索语言	( 38 )
第一节 体系分类法的原理	( 38 )
一 体系分类法的构成原理	( 38 )
二 类目的划分	( 40 )
三 类目的排列	( 45 )
四 类目名称及涵义	( 50 )
五 学科之间相互关系的处理	( 54 )
六 某些图书资料分类的特殊性	( 57 )
第二节 体系分类表	( 58 )
一 分类表的功用和结构体系	( 58 )
二 分类表的编制法	( 60 )
三 分类表的设计和编制程序	( 77 )
四 分类表的使用法	( 79 )
五 分类表的管理	( 81 )

<b>第三节 体系分类法的性能</b> .....	( 84 )
一 体系分类法的检索效率.....	( 84 )
二 体系分类法的某些问题.....	( 85 )
<b>第四章 分析-综合型分类检索语言</b> .....	( 95 )
<b>第一节 组配分类法的原理</b> .....	( 95 )
一 组配分类法的构成原理.....	( 95 )
二 “面”的划分和细分.....	( 99 )
三 类号的组配和排列 .....	(103)
四 组配分类法的类型 .....	(105)
<b>第二节 组配分类表</b> .....	(106)
一 组配分类表的结构体系 .....	(106)
二 组配分类表的编制法 .....	(107)
三 组配分类表的使用法 .....	(109)
<b>第三节 组配分类法的性能</b> .....	(111)
一 组配分类法与体系分类法的比较 .....	(111)
二 组配分类法的某些问题 .....	(112)
<b>第五章 标题词型主题检索语言</b> .....	(115)
<b>第一节 标题法的原理</b> .....	(115)
一 标题法的构成原理 .....	(115)
二 标题及其类型 .....	(118)
三 标题词的范围及选词原则 .....	(120)
四 标题词的规范 .....	(122)
五 标题词相互关系的显示 .....	(126)
<b>第二节 标题表</b> .....	(134)
一 标题表的功用和结构体系 .....	(134)
二 标题表的编制法 .....	(135)
三 标题表的使用法 .....	(136)
四 标题表的管理 .....	(144)

<b>第三节</b>	<b>叙词法的性能</b> .....	(243)
一	叙词法与单元词法、标题法、组配分类法、体系分类法的比较.....	(243)
二	叙词法的某些问题.....	(245)
<b>第八章</b>	<b>关键词型主题检索语言</b> .....	(259)
<b>第一节</b>	<b>关键词法的原理</b> .....	(259)
一	关键词法的构成原理.....	(259)
二	关键词法的表现形式.....	(260)
三	关键词的范围.....	(268)
四	关键词的规范.....	(268)
<b>第二节</b>	<b>关键词法的性能</b> .....	(269)
<b>第九章</b>	<b>情报检索语言基本方法</b> .....	(273)
<b>第一节</b>	<b>情报检索语言结构与功能的关系</b> .....	(273)
<b>第二节</b>	<b>情报检索语言为发挥“对文献的情报内容(及某些外表特征)加以标引”这项功能所采用的方法</b> .....	(275)
<b>第三节</b>	<b>情报检索语言为发挥“对内容相同及相关的情报加以集中或揭示其相关性”这项功能所采用的方法</b> .....	(282)
<b>第四节</b>	<b>情报检索语言为发挥“对大量情报加以系统化或组织化”这项功能所采用的方法</b> .....	(286)
<b>第五节</b>	<b>情报检索语言为发挥“便于将标引用语和检索用语进行相符性比较”这项功能所采用的方法</b> .....	(287)
<b>第十章</b>	<b>文献分析与标引问题</b> .....	(289)
<b>第一节</b>	<b>文献的分析与标引是影响情报处理质量的关键性工序</b> .....	(289)

<b>第三节 标题法的性能</b> .....	(146)
一 标题法与体系分类法的比较 .....	(146)
二 标题法的某些问题 .....	(149)
<b>第六章 单元词型主题检索语言</b> .....	(156)
<b>第一节 单元词法的原理</b> .....	(156)
一 单元词法的构成原理 .....	(156)
二 单元词的范围及选词原则 .....	(158)
三 单元词的规范 .....	(160)
<b>第二节 单元词表</b> .....	(163)
一 单元词表的功用、结构体系和编制法 .....	(163)
二 单元词表的使用法 .....	(164)
<b>第三节 单元词法的性能</b> .....	(166)
一 单元词法与标题法、组配分类法的比较 .....	(166)
二 单元词法的某些问题 .....	(169)
<b>第七章 叙词型主题检索语言</b> .....	(172)
<b>第一节 叙词法的原理</b> .....	(172)
一 叙词法的构成原理 .....	(172)
二 叙词的范围及选词原则 .....	(176)
三 叙词的组配 .....	(180)
四 叙词的规范 .....	(189)
五 叙词相互关系的显示 .....	(199)
<b>第二节 叙词表</b> .....	(211)
一 叙词表的功用和结构体系 .....	(211)
二 主表的编制法 .....	(215)
三 辅表的编制法 .....	(222)
四 叙词表的设计和编制程序 .....	(233)
五 叙词表的使用法 .....	(236)
六 叙词表的管理 .....	(242)

一	保证分析标引质量的重要性 .....	( 289)
二	评价分析标引质量的标准 .....	(290)
<b>第二节</b>	<b>文献分析与标引的一般方法和规则</b> .....	(291)
一	分析水平与标引深度 .....	(291)
二	分析角度 .....	(292)
三	分析方法 .....	(293)
四	标引方式方法 .....	(294)
五	导致检索效率降低的分析标引误差 .....	(295)
<b>第三节</b>	<b>文献分析与标引的某些问题</b> .....	(296)
一	适度标引问题 .....	(296)
二	集中标引问题 .....	(298)
三	利用已有标引成果问题 .....	(299)
<b>第十一章</b>	<b>情报检索计算机化与情报语言学的发展</b> .....	(301)
第一节	情报检索语言的创新和改造 .....	(301)
第二节	词表、分类表的机编化和机读化 .....	(306)
第三节	文献标引和索引编制的自动化 .....	(308)
第四节	自然语言检索的实现 .....	(310)
第五节	多种语言的结合使用 .....	(312)
第六节	检索方法的进步 .....	(314)
第七节	情报语言学研究课题和研究方法的变化 .....	(317)
第八节	结束语——情报检索计算机化对情报 语言学发展的深刻影响 .....	(320)
	<b>参考文献目录</b> .....	(322)

## 第一章 绪论——情报语言学一般问题

情报语言学是研究情报检索中的语言保证问题的一门学科，是情报检索学的一个分支。

情报语言学的主要研究对象是情报检索语言，同时也研究自然语言在情报检索中的应用问题。

情报检索语言是体系分类法、组配分类法、标题法、单元词法、叙词法、关键词法以及各种代码体系和引证关系追溯法等的统称。体系分类法等都是针对文献情报内容的检索方法，更确切地说，都是表达文献主题概念和检索课题概念的概念标识系统，是情报检索的语言工具。

情报检索语言的产生可以追溯到古代（如我国汉代刘歆创制的《七略》分类体系）。目前大多数承认的体系分类法和标题法的早期典型，即杜威创制的体系分类法和克特创制的标题法，都是1876年问世的。在此后很长一段时间内，对这两种检索方法类型的研究，一直是分别进行的。后来又出现了组配分类法、单元词法、关键词法、叙词法，等等。随着图书情报事业的发展，需要对这些方法的优缺点，特别是它们的检索效率进行比较研究。结果，就发现这些方法所要解决的基本问题是相同的，它们的基本原理是一致的，它们在方法上也是互相渗透、互相吸取的，总之，发现它们之间



有共同的规律，所以，就有必要对它们进行统一研究。这样，就逐渐形成了情报语言学这门知识。

情报检索语言这一概念大约形成于60年代初或50年代末。情报语言学方面较早的系统著作大约出现在70年代初或60年代末。

表达文献主题概念和检索课题概念的语言工具是情报检索系统诸要素中与检索效率关系最密切的一个要素。情报语言学的研究目的，是为了掌握情报检索中语言工具（包括情报检索语言和自然语言）影响检索效率的规律，以便从改进和正确使用语言工具的途径来提高检索效率。

情报语言学的研究范围大体包括以下诸方面：

(1) 情报检索中语言保证原理和情报检索语言基本理论、基本方法的研究；

(2) 各种类型情报检索语言原理、性能、编制法、使用法的研究；

(3) 自然语言在情报检索中应用的研究；

(4) 文献主题特征的研究；

(5) 情报需要特征的研究；

(6) 检索方式和检索设备及其与语言工具相互关系的研究；

(7) 情报检索计算机化与情报检索语言相互关系的研究；

(8) 文献标引和情报检索一般问题的研究；

(9) 其它有关问题（如索引法、汉字排检法等）的研究。

情报语言学的研究材料主要是：

(1) 情报语言学文献（这是已有的研究成果，也反映着

当前的研究动态)；

(2) 各种分类表、词表、代码表 (这是具体的情报检索语言)；

(3) 各种检索工具和检索系统 (这是情报检索语言以及自然语言的具体应用实例)；

(4) 各种难标引的文献 (有关新事物、新学科的文献和主题复杂、特殊的文献，可以研究文献主题特征)；

(5) 各种检索课题、咨询课题 (可以研究情报需要特征)；

(6) 学科介绍、术语词典、事物名录等参考资料。

情报语言学的研究方法较常用的有：

(1) 结构功能分析法。情报检索语言功能的完善，都是通过其结构的增加和改进来达到的。所以，从宏观和微观两个层次分析解剖其结构与功能的关系，弄清某种结构产生什么功能，某种功能是由哪些种结构产生的，产生同一种功能的不同结构孰优孰劣，以及某种结构的适用范围和条件等，是极重要的一种研究方法。结构功能分析法也适用于对自然语言的研究。

(2) 历史演进研究法。情报检索语言一直在不断进化中，所以要了解它们所采用的各种原理和方法孰优孰劣，可以从历史演进过程中看那些原理和方法是如何被扬弃的。

(3) 比较研究法。将两种或多种情报检索语言进行比较，或将产生同一种功能的不同结构、方法进行比较，或将情报检索语言与自然语言进行比较，看出它们的异中之同或同中之异，看出它们各自的长处和短处，这种方法使用很广泛。

(4) 调查整理法。在情报语言学领域，有不少已经使用

的原理和方法，尚未见之于文献，或虽有文献记载但十分零散和不全面、不系统，可采用这种研究方法了解和掌握。

(5) 归纳法和演绎法。归纳法例如根据叙词的轮排能产生多向成族和多途径检索的效果，组配分类号的轮排能产生多向成族和多途径检索的效果，关键词的轮排能产生多向成族和多途径检索的效果，而推出“标识的轮排都能产生多向成族和多途径检索的效果”的结论。演绎法例如根据标识的轮排都能产生多向成族和多途径检索的效果，标题词的倒置加正装标题见倒置标题的参照相当于标识的轮排，而推出“标题词的倒置加正装标题见倒置标题的参照也能产生多向成族和多途径检索的效果”的结论。

(6) 原理或方法的移植法。例如，用自由词作补充标引原来是叙词法所用的方法，这种方法可提高检准率，并且使用比较自由，是一种好方法，那末，是否可以将这种方法（或其原理）移植到其它类型的情报检索语言中去呢？

(7) 理想语言设计法。即设想一种功能较完善的理想情报检索语言，然后努力寻找达到这些理想标准的具体方法，进行编制——试用——优化，最后成为一个可以实际应用的结构形式。

(8) 现用语言改进法。即对一种现用的情报检索语言（如《中图法》或《汉语主题词表》）进行分析，查明其存在的问题，寻求进一步完善化和合理化的方法。

(9) 数学方法和统计方法。例如，用关系矩阵编制机内主题词表，用统计论文篇名字数的方法确定汉语题内关键词索引款目的合理长度，等等。

(10) 实验方法。例如，可用实验的方法来寻找特定机检系统标引用词的理想平均数。

电子计算机提供了很大的可能性，使一些在过去看来非常复杂的结构和似乎是难以采用的方法变得“简单”了，变得很容易做到了。这一点，在情报语言学研究上是值得注意的。

情报语言学的相关学科有：

(1) 情报检索学。这是情报语言学的上属学科，情报语言学的研究对象和所要解决的问题是情报检索学的研究对象和所要解决的问题的一部分。

(2) 图书分类学。这是情报语言学的下属学科，它专门研究情报语言学的一部分研究对象。

(3) 索引法、编目法、目录学、文献检索与利用法等。这些学科的研究对象是情报语言学理论和各种语言工具的应用方面。

(4) 语言学（构词法和词组构造法以及术语学等）。与情报检索语言特别是主题检索语言有密切关系。研究自然语言在情报检索中的应用问题更离不开语言学。

(5) 逻辑学。情报检索语言的许多方面，要应用逻辑学知识。

(6) 知识分类理论（包括科学分类和事物分类）。与情报检索语言特别是分类检索语言有密切关系。

此外，在研究、创制和使用具体情报检索语言时，还必须具备有关专业的知识。

## 第二章 情报检索语言基本理论

### 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。又称情报语言、检索语言、情报存贮与检索语言、文献语言、文献工作语言、索引语言、标引语言、标引符号、标识系统，等等。

目前世界上有一、二千种情报检索语言。例如，《中国图书馆图书分类法》、《汉语主题词表》、《国际十进分类法》、《杜威十进分类法》、《NASA叙词表》等，都是情报检索语言的一个语种；对图书文献资料所标的分类号或检索词，即是情报检索语言的语词。有数以千万计的图书馆、情报机构、档案部门以及检索刊物的工作人员和读者，正在使用着这类语言。

情报检索是情报传递中的重要环节。在当代科学技术文献数量庞大，内容专深且互相渗透，而科研、生产、教学、行政部门对情报的需要十分迫切，检索课题又很专指的情况下，情报检索问题尤其重要，并变得相当复杂；为要达到较高的检索效率，就必须作出很大的努力。

情报检索语言与情报检索效率有最密切的关系。它在情报检索过程中所起的作用是极为重要的。

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找所需的情报。所以，情报的存贮和检索是两个有着密切联系的过程，其关系如图 1。

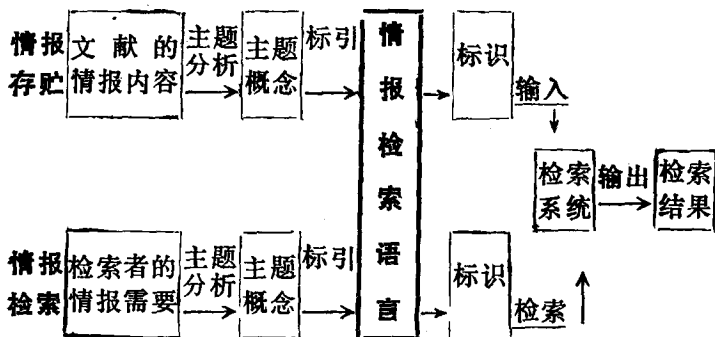


图 1 情报存贮与检索过程及情报检索语言在其中的作用

当存贮情报时，文献标引人员首先要对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，情报检索人员也首先要对检索课题进行主题分析，即把它所涉及的检索范围明确起来，使之形成若干能代表情报需要的概念，并把这些概念转换成情报检索语言的语词（标识），然后从检索工具或检索系统中查找正是用该语词标引的文献，从而找到包含有所需情报的文献。

由此可见，情报检索语言是情报检索系统的重要组成部分，在情报检索系统中起着语言保证的作用，是沟通情报的存贮和检索两个过程，标引人员和检索人员双方思想的桥

梁。如果没有情报检索语言作为标引人员和检索人员的共同语言，就很难使标引人员对文献情报内容的表达（标引用语）和检索人员对相同内容的情报需要的表达（检索用语）取得一致，情报检索也就不可能顺利实现，甚至根本不能实现。

情报检索语言作为情报检索专用语言的特点，是能简单明白又比较专指地表达文献及检索课题的主题概念，容易将概念进行系统排列，在检索时便于将标引用语和检索用语进行相符性比较；语词与概念一一对应，排除了多词一义、一词多义和词义含糊的现象，并且还能显示出概念之间的相互关系，从而能帮助情报检索人员又全、又准、又快地检索到含有所需情报的文献。

情报检索语言就其实质来说是表达一系列概括文献情报内容的概念及其相互关系的概念标识系统。它可以是从自然语言中精选出来并加以规范化的一套词汇，可以是代表某种分类体系的一套分类号码，也可以是代表某一类事物的某一方面特征的一套代码（如化合物的各种代码），用以对文献内容和情报需要进行主题标引、特征描述或逻辑分类。

情报检索语言由词汇和语法组成。词汇是指登录在分类表、词表中的全部标识，一个标识（分类号、检索词、代码）就是它的一个语词，而分类表、词表则是它的词典；语法是指如何创造和运用那些标识（单个标识或几个标识的组合）来正确表达文献内容和情报需要，以有效地实现情报检索的一整套规则。

情报检索语言的基本功能大致可归纳为如下四点：

（1）对文献的情报内容（及某些外表特征）加以标引；

(2) 对内容相同及相关的情报加以集中或揭示其相关性；

(3) 对大量情报加以系统化或组织化；

(4) 便于将标引用语和检索用语进行相符性比较。

任何一种情报检索语言，都要采取一定的方法和手段，将它的各种要素按照一定的结构结合成一个有机的统一体，以便在图书馆工作、情报工作、档案工作中发挥这四项功能，并不断改进它的结构，以发挥最佳功能。

情报检索语言的质量高低及其使用正确与否，对情报检索效率有重大影响。所以，如何提高检索效率，特别是提高检全率和检准率以及情报检索计算机化程度，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

检索效率是指全、准、快、便、省（检全率、检准率、检索速度、检索方便性、检索成本与效益），最主要的是全和准。

决定检索效率的有四个方面的因素：（1）情报检索语言的质量；（2）标引质量；（3）检索质量；（4）其它方面的因素。其中，情报检索语言的质量与检索效率关系最密切。

就情报检索语言而言，也是有许多影响检索效率的因素，所以对它必须提出多方面的要求：

1. 保证较高的检全率和检准率。或者说，漏检率和误检率要能控制在允许的范围内。

最高的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。既然检全率一般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，或者说力求降低漏检率。

最高的检准率是指检出的文献全部与情报需要有关。事实上，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。要完全避免带出无关文献也是很难达到的，所以检准率也只能达到某个百分比。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与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劳动的增加和时间的浪费，虽不如遗漏重要情报那样严重，但也会影响检索效率，所以在情报检索中，总是力求提高检准率，或者说力求降低误检率。

关于检全率、检准率、漏检率、误检率的概念见下面的图示及公式（公式有助于明确概念，虽然可以应用于评价试验，但实际很难进行十分精确的计算）：