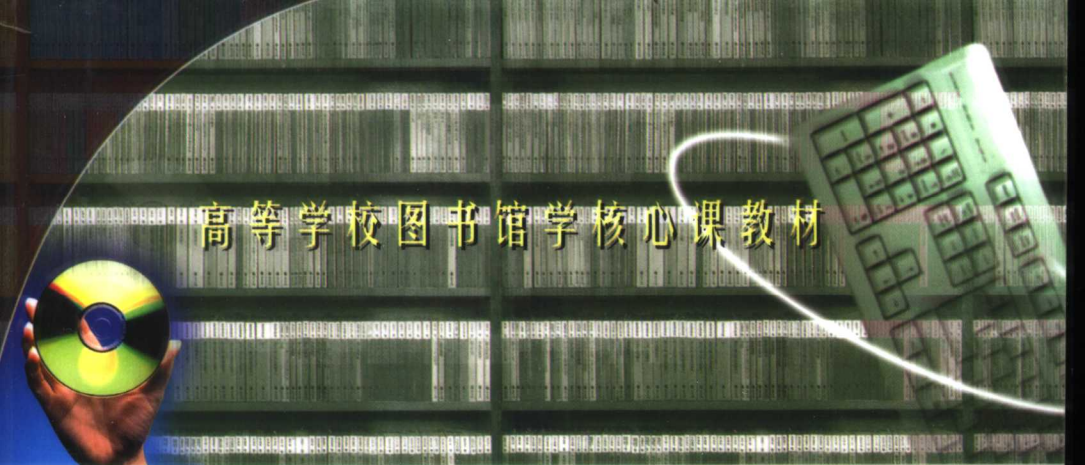




高等学校图书馆学核心课教材

信息组织的主题语言

张燕飞 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

高等学校图书馆学核心课教材

信息组织的主题语言

张燕飞 编著

武汉大学出版社

图书在版编目(CIP)数据

信息组织的主题语言/张燕飞编著. —武汉: 武汉大学出版社,
2005. 11

高等学校图书馆学核心课教材

ISBN 7-307-04731-4

I. 信… II. 张… III. 主题检索语言—高等学校—教材
IV. G254. 2

中国版本图书馆 CIP 数据核字(2005)第 118025 号

责任编辑: 严红 刘孝平 责任校对: 程小宜 版式设计: 支笛

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: wdp4@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 武汉市新华印务有限责任公司

开本: 880×1230 1/32 印张: 10.125 字数: 278 千字 插页: 2

版次: 2005 年 11 月第 1 版 2005 年 11 月第 1 次印刷

ISBN 7-307-04731-4/G · 771 定价: 16.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有缺页、倒页、脱页等质量问题, 请
与当地图书销售部门联系调换。

内 容 提 要

本书系统地介绍了主题语言的基本理论和基本知识;结合具体的词表分别论述了标题语言、叙词语言和一体化语言的构成原理和方法;阐述了主题标引、自然语言标引和检索、网络信息组织中的主题语言以及主题语言的发展等。

本书可作为高等学校图书馆学、情报学、档案学专业本科生教材,对于各级各类型图书馆、信息机构、档案部门的工作人员也有重要的参考价值。

前 言

主题语言和分类语言是构成情报检索语言的两大语系，是信息组织与检索的重要工具和语言保证。在我国，分类语言虽然有着悠久的历史，理论和实践的发展都非常完善，在长期的信息组织实践中占据主导地位。但是，自从主题语言在我国得以发展后，图书馆和信息机构纷纷利用主题语言标引文献、编制检索刊物、建立检索系统，高等学校的图书馆学、情报学、档案学专业也将其视为重要的教学内容，与分类语言形成并存格局，分别设置为两门专业基础课程。

随着信息技术的广泛应用，主题语言又有了进一步发展。为了适应科技进步和社会发展，更好地满足教学与改革的需要，系统总结在新信息环境下主题语言的理论与实践问题，更新和丰富教学内容，被提到议事日程。结合武汉大学图书馆学专业课程设置的实际，在已编写出版《文献分类学》教材的基础上，我们拟编写《信息组织的主题语言》这部教材，并得到武汉大学“十五”规划教材建设专项基金资助。

全书共8章，由张燕飞拟定写作大纲。其中第一、二、三、五章和第四章第一、二节由张燕飞编写；第六、七章和第四章第三节由邹瑾编写；第八章由张金凤编写。全书由张燕飞统稿。

本书在编写过程中，引用、参考了国内外专家、学

者的论著和论文，特别是张琪玉教授编著的《情报语言学基础》（增订二版）、侯汉清和马张华教授主编的《主题法导论》、戴维民教授主编的《信息组织》、刘湘生和汪东波研究员主编的《文献标引工作》等。在此，谨向作者表示衷心感谢。本书的出版得到武汉大学信息管理学院、武汉大学教务部和武汉大学出版社领导的关心和支持，副编审严红同志为本书的编辑出版付出了辛勤劳动，并提出了许多宝贵意见。在此，谨向这些同志表示真诚的谢意。

由于作者的学识、水平有限，加之成书仓促，书中不妥乃至错误之处在所难免，恳请专家、同行及广大读者批评指正。

张燕飞

2005年7月23日于武汉大学

目 录

第一章 主题语言概述	1
第一节 主题语言的含义和分类	1
一 主题语言的含义	1
二 主题语言的分类	2
第二节 主题语言的特征	6
一 主题语言的共同特征	6
二 各主题语言的基本特征	8
三 主题语言的作用	8
第三节 主题语言的理论基础	9
一 主题语言的概念逻辑基础	10
二 主题语言的术语学基础	14
三 主题语言的知识分类基础	19
四 主题语言的系统科学基础	21
第四节 主题语言的词汇控制	24
一 主题语言词汇控制的要求	24
二 主题语言词汇控制的内容	25
三 句法控制	29
 第二章 标题语言	 36
第一节 标题语言的原理	36
一 标题及标题的类型	36
二 标题词的选择与规范	40

三 标题的性质和标题语言的特点	45
第二节 标题表	47
一 标题表的作用和结构	47
二 标题表的编制法	48
三 美国《国会图书馆标题表》简介	49
第三节 标题表的使用	56
一 标题的选择	57
二 标题复分	58
三 标引深度和专指度确定	62
四 标题的变动与主题规范档建设	63
 第三章 叙词语言	65
第一节 叙词语言的原理	65
一 叙词语言原理和方法的综合性	65
二 叙词语言的基本原理	67
三 叙词语言的特点	72
第二节 叙词语言的词汇控制	73
一 词类控制	73
二 词形控制	74
三 词义控制	77
四 词间关系控制	78
五 先组度控制	86
第三节 叙词表的结构	87
一 叙词表的宏观结构	88
二 叙词表的微观结构	95
第四节 《汉语主题词表》简介	99
一 《汉语主题词表》概览	99
二 主表	100
三 附表	101
四 辅助索引	102

第五节 叙词表的编制与管理	107
一 叙词表的编制	107
二 叙词表的管理维护	110
三 计算机在叙词表编制与管理中的应用	111
第六节 叙词表的评价	113
一 叙词表评价的意义	113
二 叙词表评价的内容	114
三 叙词表评价的方法	115
第四章 一体化检索语言	118
第一节 检索语言的兼容互换	118
一 兼容互换研究概况	118
二 兼容转换的模式和实现手段	122
第二节 分类主题一体化语言	128
一 分类主题一体化语言概况	128
二 分面叙词表	133
三 《中国分类主题词表》	140
第三节 受控语言与自然语言的结合	152
一 受控语言与自然语言结合的必要性	152
二 受控语言与自然语言结合的原理分析	156
三 受控语言与自然语言结合及优化形式	160
第五章 主题标引	167
第一节 主题标引概述	167
一 主题标引的含义	167
二 主题标引方式	168
三 主题标引程序	170
四 主题标引的质量控制	173
第二节 主题标引方法	177
一 主题分析	177

二 主题概念转换	184
三 标识的确定	187
第三节 主题标引规则	188
一 主题标引的基本规则	188
二 各类型主题文献的主题标引	192
三 各类型文献标引规则	200
第四节 主题检索工具	207
一 手检系统中主题目录索引的编制	207
二 机检系统的文献主题标引与索引文档组织	213
 第六章 自然语言标引与检索	220
第一节 自然语言概述	220
一 受控语言与自然语言	220
二 自然语言在信息标引和检索中的应用方式	221
三 汉语自然语言检索的分词技术	223
第二节 关键词检索	226
一 关键词概述	226
二 关键词语言的类型	226
三 关键词索引的编制步骤	229
四 关键词检索的性能	230
第三节 文本检索	230
一 文本检索的特点	231
二 汉语全文文本检索	232
三 单汉字全文检索	233
四 单汉字检索的后控制技术	234
五 网络信息组织中的全文检索技术	235
第四节 后控制检索	236
一 自然语言检索与后控制检索	236
二 后控词表的功能及特点	239
三 后控词表的编制及实现	240

四 后控制技术研究及应用现状	242
第七章 网络信息组织中的主题语言	247
第一节 网络信息及网络信息组织概述	247
一 网络信息的类型及特点	247
二 网络信息用户及其特点	250
三 网络信息的组织方法	251
第二节 网络信息组织的关键词法	256
一 网络信息组织中关键词法概述	256
二 搜索引擎及其关键词检索	257
三 联机数据库中的关键词检索	260
第三节 主题词表在网络信息组织中的应用	263
一 网络环境下主题词表的类型	263
二 联机显示的网络化受控主题词表	264
三 与检索数据库结合使用的受控主题词表	266
第四节 网络信息组织中使用的其他词表	268
一 禁用词表	268
二 同义字词典和反义字词典	270
第八章 国内外主题语言的发展	274
第一节 国外主题语言的发展	274
一 以标题词语言为主的发展时期	274
二 叙词语言的发展	278
三 一体化词表的编制与发展	282
第二节 国内主题语言的发展	286
一 标题表的编制及应用	286
二 叙词语言的研究及发展	288
三 国内主题语言发展特点	295
第三节 主题语言的发展趋势	296
一 主题语言趋向标准化、兼容化	296

二 主题语言管理计算机技术化	297
三 分类语言、主题语言和自然语言将一体化	299
四 主题语言和自然语言检索技术易用性增强	300
五 自动标引技术不断完善	301
六 自然语言检索技术广泛应用并不断优化	302
七 规范化的受控语言继续存在和不断发展	303
主要参考文献	305

第一章

主题语言概述

进行信息检索，需要有信息检索系统。无论是建立手工检索系统还是计算机检索系统，都必须对大量的文献信息进行加工处理，用一种具有较小的容量、准确的语义概念的检索标识来表达文献信息的内容特征（或外表特征）。这种文献信息的检索标识需要用语言形式来表示，而主题语言便是其中的一种语言形式。本章拟就主题语言的一些基本理论及相关问题作一概述。

第一节 主题语言的含义和分类

一、主题语言的含义

为了便于检索，文献通常要用具体的检索标识来表达其特征，所谓检索标识，是以简练的语言形式表示文献信息特征，目的是用以区分和辨识文献内容，作为有序存储和检索文献信息的依据。这种特殊的语言就称为情报检索语言或称检索语言。

检索语言又称标引语言。是根据信息检索需求而创制的一种专门的人工语言。在信息组织中它用来描述文献的内容特征（或外表特征），从而形成检索标识；在信息检索中它用来描述检索提问，从而形成提问标识。

依其构成原理，检索语言可分为分类语言、主题语

言和代码语言。其中,分类语言和主题语言是两种最基本的检索语言。

我们知道,分类语言是以类目作为基本词汇,以数字或以拉丁字母和数字混合作为基本字符,采用字符直接连接的书写法,以类目的等级关系来表达复杂概念的一类检索语言,其具体体现是分类法,如《中国图书馆分类法》、《国际十进分类法》等。这种检索语言的特点是以科学分类为基础,能够比较集中体现学科的系统性,反映事物之间的学科从属关系,便于按学科门类进行浏览检索。分类法的基本体系结构是按照知识门类的逻辑序列,从总到分、从一般到特殊、从低级到高级、从简单到复杂,逐级展开的。给予文献的检索标识称为分类号。

那么,什么是主题语言呢?所谓主题语言,是以自然语言的语词为字符,以规范化或未经规范化的名词术语为基本词汇,以概念之间的形式逻辑作为语法和构词法,用语词字顺排列,主题概念以参照系统显示概念之间关系的一类检索语言。主题语言亦称主题法,它有两重含义^①:一是指具体体现主题语言的词表(标题表、元词表、叙词表等),如美国《国会图书馆标题表》、《汉语主题词表》。一是指主题整序法,即用一种语词标识处理文献,组织主题检索工具或检索系统的方法。这种方法以具有较小容量,且含义最为明确的语词来表达文献内容特征(或形式特征),并组织成主题检索系统。给予文献的检索标识是自然语言的语词。以事物为中心的直接性是主题语言的主要特点。

二、主题语言的分类

分类是区分事物的一种基本方法,它可以帮助人们准确地认识事物的基本属性特征。对主题语言作进一步分类,目的是为了更好地了解各种主题语言的特点。下面拟从不同角度来划分主题语言。

① 侯汉清,马张华.主题法导论.北京:北京大学出版社,1991

1. 依构成原理分

依构成原理,主题语言可分为标题语言、元词语言、叙词语言和关键词语言。

(1) 标题语言 亦称标题法。是一种以标题词作为主题标识,以词表预先确定的组配方式来表达主题概念的主题语言。这些用来表达主题概念的词或词组是经过规范化处理的,主要通过参照方式对词汇进行控制,并揭示标题之间的相关性,通常为比较定型的事物名称。标题法的特点是:形式直观,含义明确,操作简便。其主要不足是:概念难以多向成族,无法从多个因素、多个途径检索,灵活性较差等。

(2) 元词语言 亦称元词法或单元词法。元词语言是为了克服标题语言不足而产生的一种主题语言。它以最基本的、字面不能再分的自然语言词汇作为标识符,通过字面组配表达主题概念。元词语言的特点是:词表体积小,标引专指度高,概念可多向成族,可进行多因素和多途径组配检索,灵活性较大。不足之处是:直接性较差,采用字面组配,在字面分解与语义分解不一致时,容易产生误差,概念显示不充分,难以进行相关检索。

(3) 叙词语言 亦称叙词法或主题词法。它是以规范化的自然语言词汇作标识,主要以标识的概念组配来对信息主题进行描述的一种后组式词汇型标识系统。我国习惯称为主题词语言(主题词法),它既吸收了元词语言的一些优点,又综合了多种语言的原理和方法,成为一种具备优异性能的新的检索语言。其主要特点是:因实行概念组配原则,组配准确,标引能力强;组配方式灵活,可实现多向成族、多途径、多因素检索,检索效果较好;能适应各种检索方式和检索设备;词表结构完备,词汇控制严格,能以较少的语词表达较多、较专指、较新的概念。不足之处是:词表编制和管理难度大,对标引人员要求高,标引难度大。

(4) 关键词语言 亦称关键词法。是随着计算机的出现,为适应索引编制自动化的需求而产生的一种准主题语言。它直接用自然语言中未经控制或只作少量控制的语词表达主题概念。关键词语言

的主要优点是：标引完全专指，易于实现自动标引。不足之处是：由于对词汇不经控制或少量控制，其检索质量较差。

2. 依标识的组配特点分

依标识的组配特点，主题语言可划分为先组式主题语言、后组式主题语言和散组式主题语言三种。

(1) 先组式主题语言 先组式主题语言是指复杂的主题标识，在词表中已经组配好了的，标引时，可以直接从词表的标识中选取。标题语言就属于先组式主题语言。

(2) 后组式主题语言 后组式主题语言是指用户检索之前，主题检索系统的检索标识是单立的，并未进行组配，当用户实施检索时，才根据需要组配起来。元词语言和叙词语言就属于后组式主题语言。

(3) 散组式主题语言 散组式主题语言是指复杂的主题标识，在词表中并未组配，而是在标引阶段将表达主题概念的若干个标识组配在一起的主题语言。例如，叙词语言原本是后组式主题语言，当采用叙词表在标引阶段建立标题时，叙词语言就可当做散组式主题语言使用。

3. 依学科、专业范围划分

依学科、专业范围划分，主题语言可分为综合性主题语言、多科性主题语言和专业性主题语言。

(1) 综合性主题语言 综合性主题语言是指能供各门学科专业的文献标引和检索用的主题语言。这类主题语言因要照顾各学科专业的需求，收词一般比较全面、完备，词汇控制比较适中。如我国的《汉语主题词表》、美国的《国会图书馆标题表》就属于这一类型的主题语言。

(2) 多科性主题语言 多科性主题语言是供多个学科或某个学科群使用的主题语言。这类主题语言在尽可能照顾各学科专业需要的情况下，又具有一定的通用性。如我国的《社会科学检索词表》、日本的《日本科技情报中心叙词表》就属于这一类型主题语言。

(3) 专科性主题语言 专科性主题语言是指供某一特定学科或专业使用的主题语言。世界上大多数主题语言是专科性主题语言。如比较著名的美国《医学主题词表》就属于这一类型的主题语言。

4. 依所用自然语言语种数量划分

依所用自然语言语种数量划分, 主题语言可分为单语种语言、双语种语言和多语种语言。

(1) 单语种语言 单语种语言是指只用一种自然语言的语词为标识符所编制的主题语言。如只用汉语、英语。

(2) 双语种语言和多语种语言 双语种语言和多语种语言是指用两种或两种以上的自然语言的语词为标识符而编制的主题语言。这类主题语言有两种不同形式: 一是分别出版不同语种的版本; 一是在同一版本内采取双语种或多语种的对照形式, 如我国的《汉语主题词表》采用汉语和英语对照, 法国标准化协会的《国际技术叙词表》是法、英文对照的词表。双语种语言和多语种语言编制有利于不同语种检索语言的沟通与互换, 可促进国际间的信息交流, 尤其是网络信息的组织与检索, 迫切需要应用双语种语言和多语种语言。

5. 依使用前是否进行控制划分

依使用前是否进行控制划分, 主题语言可分为受控型主题语言和非受控型主题语言两种。

(1) 受控型主题语言是指经过规范化处理, 具有严谨的体系结构, 并对同义词、多义词等加以控制的主题语言, 如标题语言、元词语言和叙词语言等均属于这一类型。它们的共同特点是, 借助统一的检索词表, 对语词的词义、词间关系加以限定, 标引和检索时, 均须依据预先确定的检索词表对主题概念进行转换, 从而可以通过词表对文献内容的规范表达和相互关系的揭示来改进检索效果。

(2) 非受控型主题语言 亦即自然语言检索系统, 是指直接使用自然语言语词进行标引和检索的一种主题语言。关键词语言、自然语言文本检索等就属于这一类型。其特点是: 不需要检索词