京师青年教师出版资助基金
JINGSHI QINGNIAN JIAOSHI CHUBAN ZIZHU JIJIN

BIAOYIN YUYAN HE BIAOYIN FANGFA JICHU JIAOCHENG

标引语言和标引方法基础教程

陈志新　李晓娟◎编　著

北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

京师青年基金资助项目

标引语言和标引方法基础教程

Basic Course in Indexing Languages and Methods

陈志新　李晓娟　编著

北京师范大学出版社

图书在版编目(CIP)数据

标引语言和标引方法基础教程/陈志新　李晓娟编著．—北京：北京师范大学出版社，2010.10
ISBN 978-7-303-115532-2

Ⅰ.①标…　Ⅱ.①陈…　②李…　Ⅲ.①检索语言－教材　②分类标引－教材　③主题标引－教材　Ⅳ.①G254

中国版本图书馆 CIP 数据核字(2010)第 183624 号

出版发行：北京师范大学出版社　www.bnup.com.cn
　　北京新街口外大街 19 号
　　邮政编码：100875
印　　刷：
经　　销：全国新华书店
开　　本：148 mm×210 mm
印　　张：14.5
字　　数：330 千字
版　　次：2010 年 10 月第 1 版
印　　次：2010 年 10 月第 1 次印刷
定　　价：35.00 元

策划编辑：韦燕春　　责任编辑：韦燕春
美术编辑：毛　佳　　装帧设计：毛　佳
责任校对：李　菡　　责任印制：李　丽

内容提要

本教材重点研究信息组织领域的标引语言和标引方法问题。首先，分析索引语言和索引方法的理论基础；其次，重点讨论分类的标引语言和主题的标引语言，并通过大量的实例增强对标引语言和标引方法的理解；再次，以《中国分类主题词表》为重点，讨论标引语言的编制、评价和兼容问题，从整体上研究标引语言和标引方法；最后，作为实践性比较强的部分，介绍和研究引证关系情报检索语言、后控词表、信息挖掘、单汉字标引、自动文摘和自动标引等标引方法。

目　录

第一篇　引论

第二篇　标引语言

第三篇　标引语言的编制、评价和兼容

第四篇　标引方法

第一篇　引　论

虽然标引语言和标引方法的研究是多方面的，但是可以用几个中心内容把这个领域统一起来。在这里，我们所讨论问题的基本思想，将会贯穿全书。同时这也有助于我们对本书产生概括性的了解。

第一章　标引语言和标引结果

通过对标引语言和标引方法的学习，人们可以提高自己的学习效果和学习能力——这是从宏观和最终效果层面讲的。从具体的层面，标引语言究竟是做什么的呢？

标引语言是用于表达一系列文献信息内容及其相互关系的概念标识系统。人类的社会日常生活，产生大量的信息，如果任其自然状态发展，不加以有效的信息组织方面的干预，将会混乱无序；人类的智力活动，经过世世代代长期积累，形成浩瀚的知识海洋，如果不加以有效的知识管理方面的干预，也将会杂乱无章。

日常生活也好，伴随人们一生的学习活动也好，都需要我们有选择并且有效地利用已存在的信息和知识。

现在我们对周围的信息环境、知识环境还没有感到太混乱，这在很大程度上是应用了标引语言的基本原理和方法，对信息和知识进行了有效组织的结果。

在这里，我们将介绍两种典型的标引语言和方法，分别是分类法和主题法。

一、分类式的语言和方法

分类法是表示和组织知识的两种基本方法之一。

类是指具有共同属性的事物的集合，“物以类聚”是人们长期以来认识事物的一种方法。分类检索语言也称分类法，是对根据一定原则组织起来的许多类目，通过标记符号来代表各级

类目以固定其先后次序的分类体系。

知识的载体包括两种，一种是人类的大脑，另一种是计算机和文献。我们可以有效地组织和整理自己头脑中的知识，思考一下哪些方面欠缺、各种知识之间的联系和继承性，看看哪些方面需要更加努力地补充和更新，这样的工作，我们每一个人每天都在做。而对于别人头脑中的知识，似乎只有特定职业的人，比如教师和科学家，才能够帮助别人去有效地组织和整理。

标引语言，虽然对于头脑中知识的组织，能够起到一定的宏观指导作用，但是它更擅长组织存储在计算机和文献中的知识。

比如图书馆里面的书籍，假如没有工作人员的有效组织，经过几十年上百年的积累，将会变成凌乱无章的“书山”、“书海”。如果对其进行有效的组织和整理，首先会想到，用我们头脑中已经存在的关于整个人类知识的分类次序，建造一座大楼，这座大楼有几十层高，每层楼有几百个房间，每个房间又有几个或十几个小室。如同真实的房间的号码一样，我们给这座楼的每一个房间，都贴上各种知识和各门学科的标签。其次，审查“书山”里的每一本书籍，将其放到合适的房间中去。

此时，“书山”已经变为“书籍的大厦”，由于每一个房间都是按照“知识的秩序”建立的，人们当然知道到哪层楼的哪个房间中，就能找到其需要的书籍。

这就是大家已经熟悉的分类方法。

建筑这种大厦，我国已经有了比较公认和统一的标准——《中国图书馆分类法》，用这个公认的“知识的秩序”建设的“大

厦”，共有 22 层楼高，分别是：

A　马克思主义、列宁主义、毛泽东思想
B　哲学
C　社会科学
D　政治、法律
E　军事
F　经济
G　文化、科学、教育、体育
H　语言、文字
I　文学
J　艺术
K　历史、地理
N　自然科学总论
O　数理科学和化学
P　天文学、地球科学
Q　生物科学
R　医药、卫生
S　农业科学
T　工业技术
U　交通运输
V　航空、航天
X　环境科学、安全科学
Z　综合性图书

每一层楼，包括许多房间，比如，其中第六层楼“F 经济”包括 9 个子类(房间)：

(1)经济学
(2)世界各国经济概况、经济史、经济地理

(3)经济计划与管理

(4)农业经济

(5)工业经济

(6)交通运输经济

(7)邮电经济

(8)贸易经济

(9)财政、金融

当然，每一个房间，又包括很多室。这样一级一级地走下去，这个大厦一共22层楼，共1 500个房间，30 000～40 000个小室。我们国家的知识系统，主要是使用这个体系完成信息组织工作的。

分类的方式，在国外也是最主要的知识组织方式。比如，由美国人杜威编制的《杜威十进制分类法》(*Dewey Decimal Classification*)，是一部在国际上出现最早、流行最广、影响最大的图书馆分类法。该分类法共10层楼，每层10个房间，每个房间10个小室，形式上比我国的分类体系规整，形成10层楼、100个房间、1 000个小室的格局。

10层楼的情况如下：

000 Computers, information & general reference(总论)

100 Philosophy & psychology(哲学和心理学)

200 Religion(宗教)

300 Social sciences(社会科学)

400 Language(语言学)

500 Science(科学)

600 Technology(技术科学)

700 Arts & recreation(美术)

800 Literature(文学)

900 History & geography(历史和地理)

每一层楼包括 10 个房间，比如第 7 层“600 Technology (技术科学)”包括如下 10 个房间：

600 Technology(技术科学总论)

610 Medicine(医学)

620 Engineering(工程学)

630 Agriculture(农业)

640 Home & family management(轻工业和手工业)

650 Management & public relations(管理和公共关系)

660 Chemical engineering(化学工程)

670 Manufacturing(制造业)

680 Manufacturing specific products(专门产品的制造)

690 Building & construction(建筑业)

这样一级一级下去，美国的知识组织大厦，共 10 层楼高，100 个房间，1 000 个小室，整个知识组织的大厦共包括大约 30 000 个基本单元(approximately 30 000 numbered concept definitions in the Dewey schedules)。由此看来，无论中国还是外国，组成知识分类大厦最小单位的数目是不约而同的。

总之，用学科分类知识，建造起我们容易理解和掌握的体系，用来处理纷繁复杂、数量浩大的知识载体，把无序的知识载体变得有序化，我们称其为分类式的语言和方法。

二、主题式的语言和方法

主题法是表示和组织知识的两种基本方法之一。

分类法在被组织的知识之外，重新建立一套体系，借助新

建体系的逻辑次序使被组织的信息有序化。

主题法与分类法走的是完全不同的道路，认为被组织知识的内部，已经包含一种能够自我组织的力量。比如学科概念、名词术语、理论学说以及大量的事物名称，可以直接用来组织知识。

拿经济学来说，列举几百个甚至几千个经济学的名词、术语、经济理论学说和经济学家的名称，假如其中有“商品流通、宏观经济学、微观经济学、货币、古典经济学、亚当·斯密(Smith Adam)”这样几个词汇，某书籍只讨论商品流通，我们用“商品流通”这个词汇表示并代替它；另一本书不仅讨论商品流通而且讨论宏观经济学，我们用“商品流通和宏观经济学”两个词汇表示并代替它……此时，建立了书籍与词汇的对应关系，假如这些词汇建立了次序，相应的词汇所代表的书籍也就随之有序化了。很显然，词汇可以像字典那样，依据拼音或字形的次序排列起来，那些书籍也随之建立了秩序。

我们把用简练和具有代表性的词语直呼其名的信息组织方法称作主题法。

我国综合性的主题词表是《汉语主题词表》，它收录的词汇是《中国图书馆图书分类法》的 3 倍，大约 10 万个主题词，很显然，它的具体性和专指性更好。

三、分类式语言与主题式语言的区别

分类式语言和主题式语言，作为两种基本的信息组织方法，在信息组织过程中各自发挥着不同的作用，两种语言的侧重点不同，分类式语言更强调学科的角度，主题式语言更强调事物名称的角度。用分类法组织，能够集中在一起的信息资

源，如果以主题法组织很可能分散开；而以主题法组织，能够集中在一起的信息资源，如果用分类法组织也可能分散开。我们把这种现象称作“集中与分散的矛盾”。

谈到主题法与分类法的区别以及“集中与分散的矛盾”这个问题，下面是我国的一个经典例子：

用主题法组织，从各个方面讨论“茶叶”的文献，能够集中在一起。但是，改用分类法组织，同样讨论茶叶的文献，按照不同学科，则被分散到各个领域中：

茶叶贸易　F 经济类的贸易经济小类
茶叶栽培　S 农业类的农作物栽培养植小类
茶叶加工　T 工业类的轻工业和手工业小类

国外在谈到主题法与分类法的区别以及“集中与分散的矛盾”的问题时，也有一个非常经典的例子：

不管从哪个方面讨论“衣服”的文献，在主题法这里，均被集中在一起。但是，用分类法组织，将被分散开来：

衣服对人心理状态的影响	B 心理学
衣服的穿戴习惯	G 文化风俗
服装设计	J 艺术类的设计艺术

这种区分，对我们的检索实践非常有帮助，应该在检索过程中学会切换检索方法。因为某些检索需求，用分类方式比较好，用主题方式效果不好；另外的情况可能相反，主题方式比较好，用分类的方式效果不好。

例如，检索可以划分成四种主要类型：

(1)检索某一事物某一方面：

长江(抗洪)。

(2)检索某一事物全部有关的文献：

长江(抗洪、发电、航运、养殖)。

(3)检索许多事物同一方面：

(长江、黄河、嫩江)抗洪。

(4)检索一类事物的全部方面的有关文献：

长江(抗洪、发电、航运、养殖)。

黄河(抗洪、发电、航运、养殖)。

嫩江(抗洪、发电、航运、养殖)。

很显然，第(1)、第(2)和第(4)种检索需要，使用主题法比较好；第(3)种检索需要，使用分类法的效果会更好。适当地选择检索方法，将在信息检索中为我们带来很大的便利。

四、标引、标引结果与检索的关系

对于知识表示和知识组织领域，分类法和主题法是两种基本方法。运用这两种方法，处理原始信息，形成标引结果的过程，称为标引。

从专业角度理解，“原始信息”包括两个方面，一方面，指以文献为主要代表的知识产品。比如，清华大学出版社在1991年出版了谭浩强编著的《C语言设计》，此书320页，共13章内容。另一方面，指以利用文献为主要获取知识手段的用户的知识需求。比如，像“下学期开设C语言课程，找一本合适的书看看吧”之类的需求。

“处理”指使用分类法或主题法，对“原始信息”的两个方面都进行加工和替代。清华大学的书籍，我们用分类的方法，将其定位在“工业技术——计算机技术——计算机程序设计语言——C语言”上，赋给“TP312C”的代号代表该文献；用户的

知识需求，有时比较模糊，有时是一种潜在的心理需求，必须转换成为明确的需求，使用分类的方法，同样将其定位在“工业技术——计算机技术——计算机程序设计语言——C 语言”上，同样赋给“TP312C”的代号，代表用户的需求。

“TP312C”是对原始信息的两个方面，使用一定的标引方法处理后形成的结果，称为标引结果。

所以，标引包括对信息和信息需求两个方面的处理。

理解标引概念时，一位学生甚至用起了古文：“标引者，标示指引也。于纷繁之大千文献中，抽取线索，理出头绪，物归其类，概繁为简，化面为点，以明确浅显之法，显见易知之章法头绪，标示种种知识，分门别类，作为对文献用户的标示指引。”

检索是形成信息需求并满足信息需求的过程。对检索的一种理解是，从信息集合中，找出含有特定内容信息的过程。这个定义，比较强调信息集合的复杂性、庞大性，以及用户信息需求的独特性、单一性，是从结果的角度下定义。

从过程的角度，检索是对信息产品的标引结果和信息需求的标引结果的匹配以及信息产品的调出。一个完整的流程是：用户→信息需求→对信息需求内容分析→用标引语言转换成标引结果(信息需求的转换结果)→满足信息需求的转换结果(信息源的转换结果)→信息源。匹配是对信息产品的标引结果即信息源→选择信息→对信息内容分析→用标引语言转换成标引结果，与信息需求的标引结果即用户→信息需求→对信息需求内容分析→用标引语言转换成标引结果进行比对的过程。

一旦匹配成功，说明信息源中一定含有用户需要的信息，