

主 題 法

原 理 与 实 践

北京师范大学图书馆学系

一九八五年二月

第二章 主题标引工作

第一节 主题标引的意义与工作程序

一、主题标引的意义

文献的主题标引，“是对文献进行主题分析，从自然语言转换成规范化的检索语言的过程，即对主题分析结果赋予检索标识的过程。”“给予主题词标识的过程，称为主题标引。”※

文献的主题标引，就是利用一种特定的检索语言，揭示（描述）文献的内容，以便提供检索利用。在文献的整理加工过程中，主题标引工作的重要性在于：

① 通过主题标引将入藏的文献中，对文献利用者有用的重要信息给以主题标识，提供多种检索途径，使文献得到充分利用；

② 通过主题标引将同一主题的文献加以集中，便于比较全面地查找到所需课题的文献；

③ 只有通过主题标引，才能构成主题字顺检索系统，将文献内容编排成系列，提供检索。

综上所述，主题标引工作质量，直接关系主题字顺检索系统的建立和利用，影响文献作用的发挥。因此，主题标引是检索工作的基础，是准确、全面和高效率地进行文献检索的保证。

文献标引工作和检索工作有着密不可分的关系，它们的目的能

※引自《中华人民共和国国家标准 GB 3860—83 主题标引规则》第1页 国家标准局批准1984年9月1日实施

是为了充分利用检索系统中存储的文献；它们所使用的检索语言必须一致，从这个意义上说，描述同一主题的标引词和检索词应该是一致的；此外，标引和检索课题的输入过程也是基本上相同的，都有一个将主题转换为检索语言的过程。

但是，检索工作既有将检索课题输入检索系统的过程，也有从检索系统输出检索结果的过程。而且在文献检索服务中，检索工作属于直接面向用户的服务工作，检索课题具体，服务对象明确，服务效果显著。

文献标引工作则不包括输入过程，与服务对象不直接发生关系。服务效果也不显著。但事实上，文献标引恰当与否，直接影响检索系统和检索工作的质量。例如：

标引时由于标引人员不适当地扩大了标引的范围，标引的主题词或者与检索系统的性质、服务对象无关，或者对于用户没有参考价值，与文献论述的问题无关，就会造成过度标引。其结果，不但浪费人力、物力和空间，而且会消耗不必要的查找时间，影响检索效率，以及由于提供了毫无参考价值的信息，降低文献检索的查准率，影响了检查效果。

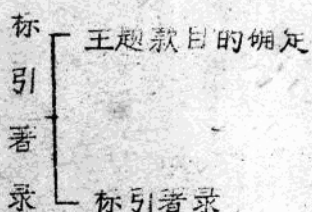
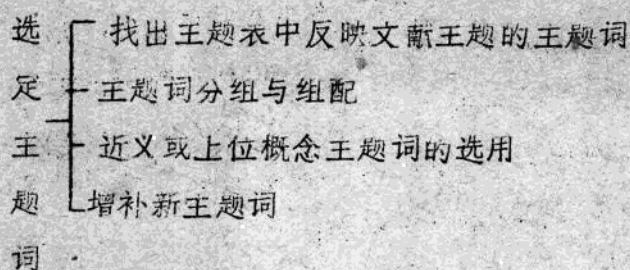
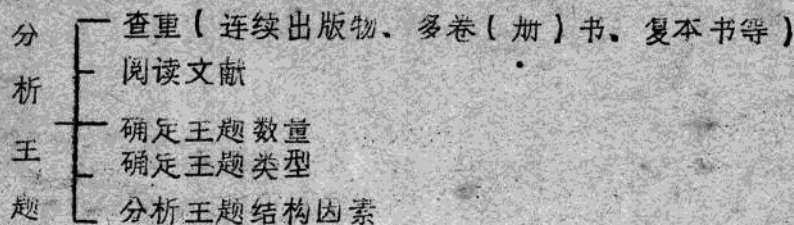
与过度标引相反，如果标引过粗，主题词概念大于文献主题的概念，则主题词包含的文献复杂，专指性差，同样会耗费不必要的查找时间和降低文献检索的查准率。

如果标引人员漏标文献有参考价值主题，那末，就会使这篇文献被埋没，从而影响检索系统的服务效果。

二、主题标引工作过程

主题标引工作过程与分类标引工作过程基本相似，包括分析主

题, 选定主题词, 给标识三个阶段。(·见下图)



由于主题表类型不同, 主题标引的繁简也各不相同。一般来说, 使用标题表较之使用后组式的叙词表进行标引, 就不必考虑主题词的组配和主题款目的拟定问题, 相对来说比较容易一些。但其他步骤基本上一样。

三. 对主题标引人员的要求

主题标引人员的素质, 决定标引工作的水平。主题标引工作技术性很强, 即使是多年从事分类工作的人员, 也要经过系统的培训,

才能胜任工作。一般来说,主题标引工作应由馆员以上水平的人员担任,尤其是今天,各行各业强调科学管理,图书馆情报工作各部门都纷纷制订工作规范,对于各项工作的工作内容,工作量,工作要求均作出规定,对于主题标引工作者来说,要求是比较高的,他必须能根据文献内容及规定的主题表,严格按照标引规则标引出文献的主要内容,校对前错误率不能超过5%,他必须每天完成规定的工作量,一般需标引20种中文图书或15种西文图书。

为此,主题标引人员必须具备的基本要求:

1. 掌握本部门规定使用的主题表,了解主题表的结构,范围,描述主题的方式特点,以及使用的规则、方法和要求。

同时,要善于利用他人标引的成果,对与本部门有关的主题表要有所了解。

2. 了解本部门检索系统的性质,特点,对文献标引的要求。

同时,要了解用户的需求,包括他们的职业、工作性质,专业水平,查找的习惯等。

3. 必须善于对文献进行分析综合,能准确表达文献主题,具有较高的语言表达能力。

4. 必须具有一定的专业知识,并能了解科学技术发展动向,具有处理各门类知识的文献的能力。

为了适应文献检索工作的发展,除了在正规大学培养主题标引人员外,应大力进行在职人员的培训,争取今后几年内形成一支具有一定水平的主题标引人员的队伍。

第二节 主题标引的规则与方法

一、主题标引的规则

主题标引规则的制订,是为了保证文献主题标引的一致性和准确性,提高标引工作的质量和检索效率。

主题标引规则往往是使用某种类型或具体的主题表的規定,或者某一部门,某一单位主题标引工作的規定。例如,《中华人民共和国国家标准 GB 3860-83 文献主题标引规则》是“适用于依据各种类型的汉语主题词表进行文献主题标引。”(引自该标准),《北京图书馆中文图书主题标引工作条例》“主要是作为北京图书馆中文图书铅印卡片,头行主题标引的依据。同时也是作为我馆使用计算机生产中文机器可读目录和编制国家书目,提供主题检索途径的基本依据。”(引自该条例)。关于《汉语主题词表》的使用规则,科技文献出版社1979年出版了《〈汉语主题词表〉使用方法说明》,北京图书馆1981年出版油印本《〈汉语主题词表〉手工检索标引工作手册(初稿)》。

进行主题标引工作主要遵循具体的主题表或具体单位的标引规则,但也要遵循一般的主题标引规则。

一般的主题标引规则的内容包括:

1. 主题标引的指导思想。一般要求主题标引人员对文献论述和研究的事物,对象和问题应客观反映,不能掺杂个人的观点进行评定或扬弃。

2. 关于标引过程中各步骤的要求,即对主题分析,选定主题词各步骤的要求。其中对主题标引中的标引深度(对某一文献所给予的主题词的数量),不同的检索系统应根据实际需要(检索系统

规模、检索方式、文献类型等)作出具体规定。

3 文献主题的表达方式。包括主题词的来源(自然语言还是规范化语言),主题词的组配,标题的结构等,应作出相应规定。

标引人员在从事标引工作之前,必须熟悉标引规则,按照规定进行标引。

二、主题分析的方法

主题分析是主题标引的第一个步骤,它的工作过程包括浏览文献,确定主题的数量和类型,分析主题的结构因素等。对于多卷书,连续出版物,上、下册分别出版的图书,需要首先查重,了解以前标引的情况。

主题分析是紧张的脑力劳动,它要求标引人员在规定的时间内,正确地判断出文献的主题,并从中进行筛选。主题分析的结果,直接影响整个主题标引的工作质量,以至检索效率。由于文献的主题,它所研究的问题的多样性和专业性强,文献的表达方式,作者的水平各不相同,标引人员要对文献作全面的了解,就必须有广博的学识,较深的语文程度(包括外语水平),有较强的分析综合能力。此外,还要掌握主题分析的方法。

现就主题分析过程中,有关主题数量和类型,结构等问题,进行一些探讨。

1. 主题数量的确定

无论是一篇文章,还是一套连续出版物,它们论述的问题——直接论述的问题和间接论述或涉及的问题,总是很多的。要确定主题数量的较佳方案,必须同时考虑几种因素,主要是检索系统的性质和

服务对象(手工检索还是机器检索,专业性还是综合性等);标引方式的决定;以及文献形式(图书、论文等)。

一般来说,手工检索系统(例如卡片式目录、索引),书本式检索系统(例如索引、书目、文摘的主题字顺索引),主题数量不能太多。

卡片式检索系统检索时需逐张翻阅卡片,即使有详细导卡,翻阅也不方便,而过多增加主题数量,势必增加卡片,既占用很大空间(卡片柜和目录室),也使排卡片工作更加繁重,因此,主题的平均数量以1—3较合适。

书本式检索系统检索时不必逐卡翻阅,容易过目,但主题数量增加,就会增加篇幅,印刷的成本则相应提高,因此,主题平均数量一般也应控制在2—3的范围内。

作者曾对美国国会图书馆出版的《国家联合目录》(*Library of Congress Catalogs NUC*) 1984年4月号抽样统计,1,099种图书中,一个主题的占29%,二个主题占29.8%,三个主题占21.3%,四个主题仅占1.7%。而且4—6个主题的25种书几乎都是输入MARC磁带的。

另外,据董春元对约4万篇文献的主题词分析统计,*用一个主题词能表达文献主题的,占文献总数不到5%,用两个主题词的占4.5%,用三个主题词组配的占4.0%,必须四个以上主题词组配的不超过1.0%。

以上调查统计,说明无论是图书或文章,在手工检索系统中,

*引自“索引款目中主题词数量的确定”,董春元,《情报工作通讯》,1983年第3期,第11—13页。

1—3个主题基本上可以对文献主题表达清楚。

计算机检索系统无论是文献存储能力和检索速度，均非手工检索系统可比，因此，表达文献主题的数量可增加到十个以上。

检索系统的性质和服务对象，对于主题的数量，选择的专业范围也有影响。专业性检索系统（如专业图书资料部门的目录，专门学科的主题索引等）可根据需要进行文献的部分标引（也叫重点标引），即只标引与本专业有关的文献主题。例如：

“当归对心血管系统的药理作用”一文，除了中药“当归”的药理作用这个主题外，文中还论述“当归对心血管系统的作用”，“当归对药物的疗效”等问题，读者可能是研究药物的，也可能是心血管系统疾病的医生。但是在药物专业的检索系统中，只选取“当归”这个主题。

决定文献主题数量的因素，还有文献的类型和标引的方式。

一本书或一篇文章研究讨论的问题可以归纳为一、二个概括性主题，也可以把著作中包含的有参考价值的问题一一列出。

仅就全书或整篇文章标引出主题称为整体标引，整体标引的主题数量较少。有时在整体标引的同时，根据需要对文献中重要的，参考价值较大的主题（章节或附录的主题）作补充标引，这种标引方式一般称为分析标引。整体标引多用来标引图书，特别是具体的图书馆比较适用。

为了全面反映文献主题，需要将文献中有使用价值的内容进行全面揭示，这种标引方式称为全面标引。全面标引能提高文献的利用率，提高检索效率。但是由于主题数量大大增加，必然花费更多的人力、物力，提高经济成本，检索系统的体积也相应增大。这种标引方式一般不采用。

对于汇编性文献，例如连续出版物、会议录、成果汇编以及多卷书、文集等，标引时有两种方式，一种是综合标引，以整套出版物的主题进行标引，而不考虑各分册或单篇文献的主题内容，这种标引方式只需数量很少的主题，但会将很多有用的文献埋没；另一种是分散标引，即对各分册或单篇论文分别标引，这种标引方式可使主题数量多至上百个。

2 主题的选择

主题分析实质上就是分面分析，即对文献论述的问题按不同的范畴面，分析构成主题的组成因素。选取主题是为了提供编制主题书目或主题索引的依据，因此必须考虑主题是否具有检索意义，是否符合用户的需要。

因此，主题因素根据检索功能可分为具有检索意义的因素和从属因素。具有检索意义的因素在检索系统中能作为检索款目。从属因素则不必作为检索款目，而作为检索款目的说明语。

作为从属因素一般是通用的泛义概念，例如，设计、施工、发展、应用、研究、试验等。时间、空间和文献类型在社会科学领域中，有时可作为检索款目。例如：《20000年的欧洲》、《美国参考书年度索引》(Index to American reference books annual)二书，“欧洲”和“参考书”可作为检索款目，但这些概念一般只作为从属因素，不作检索款目。

不同的学科领域的范畴面，具体内容有所不同。例如：

工程技术	{	实体——设备、装置、元件、材料……
		手段——操作、工艺、措施……
		特性——现象、性能、效应……

送字 { 实体——器官、药物、设备、病毒……
手段——预防、治疗、临床、制药……
特性——病毒、疗效、药性……

农业 { 实体——作物、牲畜、肥料、害虫、土壤……
手段——栽培、育种、饲养、保护、加工……
性能——生理、生态、作用……

社会科学 { 实体——人物、社会结构、事件、社会现象……
过程——发展过程、活动方式、心理活动……
条件——理论、立场、观点、方法……

上述范畴面的内容，大多数是文献研究的对象，因此都具有检索意义，可作检索款目。

在阅读文献的基础上，标引人员根据文献内容，从中寻求下列问题的答案作为文献的主题：

① 文献研究论述了哪些问题，其中的关键问题是什么？是否需要进行问题方面的深入？

② 文献是否讨论到特定的产品、决定性条件或最后现象？

③ 主题是否包括具有活动力的概念（例如动作、工艺操作、过程等）？

④ 主题是否受确认行动（动作）的影响？文献是否涉及这一动作的对象？

(5) 文献是否涉及完成动作的特殊方式方法(具体条件或措施),例如特种仪器(特殊设备),特种技术或方法等。

(6) 文章前后考虑的因素是否关系到特定地点或环境?

(7) 是否有确认的变量或自变量?

(8) 讨论的主题是否在一般研究领域中有无联系的特殊观点?

(9) 文献是否具有检索意义的外衣特征或类型可作主题,例如论文集、会议录等。

(10) 标引图书时,是否可利用在版编目中的主题词,以及其他主题标引的成果?

(11) 读者(检索系统的用户)可能从哪些角度来查找文献,本文献有没有这些主题因素?

文献主题的筛选工作是十分重要的,在众多主题中排除不必要的,或检索意义不大的主题,对提高标引工作质量具有重大意义。

筛选的主题包括:

(1) 仅在文献中提及但没有论述的名词术语;

(2) 使用价值不大的主题或一般消息报道;

(3) 虽有一定使用价值,但与本检索体系的专业性质、内容不相符,用户利用不多的主题;

(4) 一般性的结论、理论等。

主题筛选工作是提高检索效率,保持主题标引工作中合理的广度的重要保证。

3 主题类型的确定

主题类型的确定，与主题款目标目的构成有关。不同主题类型会产生不同的标引方式。根据文献论述问题多少，主题类型还区分为单主题和多主题两种。

单主题表示一篇文献只论述一个主题。只使用一个主题词来表达文献主题叫一维标引。一维标引的主题又称单一主题或单元主题。利用两个或两个以上主题词表达文献一个主题叫多维标引。多维标引的主题又称多元主题或复合主题。

多主题又称并列主题，表示一篇文献论述两个或两个以上主题。多主题的标引称分组标引，即将各主题以单主题的方法标引。

复合主题和多主题标引时，都需要将两个以上主题词进行组配，因此都属于组配标引。进行组配标引的复合主题中，主题词之间相互关系的反映，要按照有关规定进行。

三、主题词的选定与组配

主题词的选定，首先应确定采用哪种类型的检索标识。目前一般采用三种检索标识进行标引：规范化语言，自然语言（又称自由词），以及上述两种语言的混合使用。

采用规范化语言作标识称为受控标引，标引时必须使用编制好的标题表或叙词表作为标引工具。将主题分析后选取的主题，从词表中找出最专指的主题词。例如，《机械制造文摘》、《中国药字文摘》及中外文医学科技资料目录等。

采用自然语言作标识称为自由标引，标引用的自由词或关键词一般取自文献中用以表达文献主题的名词术语，例如，我国1982年出版的农字文摘。

目前也有部分检索刊物采用混合标引（半控标引）方式，一般

~ 12 ~

以规范化语言作检索款目，自然语言作说明词语。

采用规范化语言作为主题词时，必须将文献主题转换成主题词。转换过程进行得快慢，主要决定于以下因素：

① 标引人员对描述主题的术语（包括同义词、近义词），以及它们的从属关系的了解程度，即对文献论述的学科内容的熟悉程度。

② 标引人员对主题表的熟悉程度。

③ 标引人员对主题表组配方式的掌握。

④ 主题表的编制质量。

在使用主题表时，如果有范畴索引，文献主题比较简单，可利用范畴索引选定主题词，但为了准确起见，特别是确定复合概念的主题词，应利用字顺表，并注意表中的参照项。尤其是《汉语主题词表》的社会科学部分，有相当多的复合概念使用一个复合主题词表述，例如：“会计标准化”、“商业管理”、“教育理论”等。

另外，有部分主题词要通过参照关系进行分析比较才能正确选定。

标引工作的专指性与选定主题词有密切关系，主题词选用不当，不能确切表述文献主题，会造成检索误差。因此选词时，既要避免选用泛指的主题词代替更专指的主题词（即所谓以粗代细），也要避免选用下位概念的主题词，代替切合文献主题的上位概念主题词（即所谓以细代粗）。

当主题表中没有专指的主题词能直接表述文献内容时，为了提高描述主题的专指性，达到良好的检索效果，经常用几个主题词组配成一个完整的概念。

从文献使用主题表进行标引时，字顺表中基本上已固定组配的形式，

而且标引的规则较严格。例如《美国国会图书馆标题表》(LCSH),《医学主题词注释字顺表(英汉对照)》等。对于标引人员和检索人员来说,使用比较方便,容易在选定主题词时取得一致。

使用叙词表进行标引时,则需要标引人员或检索人员利用表中的主题词进行组配。字顺表中一般不指示组配的原则与方法。

主题词的组配,实质上是词义与语法的结合,使相互之间单独存在的主题词发生逻辑上的联系。

主题词的组配不是简单地从字面上将几个主题词相加,而是将表示若干个分概念的主题词组配起来,表示一个完整的主题概念。

例如,

全天候轰炸机应是“全天候飞机”+“轰炸机”

不应是“全天候”+“轰炸机”

主题词组配所表示的语法关系,一般是交叉关系,从属关系,限定关系,应用关系等。语法关系的表达方式,一般利用主题词的词序先后和使用组配符号。

词序先后包括正装和倒装,正装形式一般用横符“—”表示“—”之前为表示主题因素的主题词,用作检索款目,“—”之后为说明词语。例如:

“气管切开的副作用” 标引形式为:

气管切开—副作用

倒装形式一般用逗点“,”表示,逗点之前仍为检索款目,表达主题的词前后倒置。例如:

“福建地区稻瘟病的研究” 标引形式为:

稻瘟病, 福建

其他组配符号的使用，各主题表有具体的规定，《汉语主题词表》规定用圆括号“（）”表示概念的限定关系，冒号“：”表示概念的交叉关系。概念的影响关系则采用以下分式：

主标题——关系，对象

主标题——作用，对象

例如：

“辩证法与形而上学的关系”可标引为：

辩证法——关系，与形而上学的

标引人员必须熟记规定的组配符号，才能清楚地、确切地组配出一个完整的主题。

在选定主题词过程中，由于科学技术的发展，主题表中的主题表往往包括不了新的文献主题。标引人员可根据实际需要增补新的主题词，增补新词应填写“主题词增删改记录卡”，同时要考虑词的参照关系。应在一定的条件下增补新词；其中最主要的是新词是否具有实际检索意义，是否有相当的文献量。另外，现有的主题词或主题词组配后是否仍不能表达文献的专指概念等。

四、 标引著录

标引著录是标引工作的最后阶段，首先应确定采用组配符号和表达主题的形式，目前国内检索系统中，表达主题主要有两种形式。

第一种形式是单词型，即只用一个主题词作主题款目。《中国药学文摘》，《机械制造文摘（机床与工具分册）》，《国外电子科技文摘》的主题索引属于这一类型。其标引形式举例如下：

文 献 号	篇 名	主题款目
	黄柏 的组织培养	组织培养
	当归对心血管系统的药理作用	当归
8212041	粉末冶金高速钢拉刀	拉刀
8212321	数字直流伺服驱动在电火花线 切削机床上的应用	电火花线 切削机床
19251(81)	由弥散产生的内应力玻璃断裂 韧性的影响	玻璃
14851(81)	在电话局环境中测定的有机 物质	电话局

第二种形式是传统标题法型，即每一主题款目用一个或二~三个主题词，采用“主题——副标题”的著录形式。例如《中文科技资料目录——医学》、《国外科技资料目录——医学》，以及正在开展的中文图书主题标引的形式。其标引形式举例如下：

文 献 号	篇 名	主题款目
YY8301401	白喉喉梗阻并自发性气 胸的探讨	白喉—并发症 喉症病—病因学 喉症病—治疗 气管切开—副作用 气胸—病因学
Y Y8302401	矽肺眼底改变的初步 报告	矽肺—病理生理学 眼底—病理生理学