

写在前面

课程名称：情报语言学

参考教材：《文献分类法主题法导论》马张华 侯汉清 薛春香 编著

授课教师：何琳老师

文档中的内容是根据老师 ppt 的框架和内容整理而成的，部分 ppt 上没有说清楚的内容摘取了参考教材上的部分描述。题型为老师上课讲述，根据自己对课程的理解猜测每个题型可能考的内容，仅供参考；论述题由于老师出题脑洞太大，实在猜不到，考场上结合结合自己脑中所有的相关知识尽力答吧。题目为考试后人脑记忆并复述，大体意思相同，表达可能略有出入。

文档中重点内容用✖标出，次重点内容用☆标出，同样根据老师的讲述和自己的猜测，每届重点可能不相同，也仅供参考。重点内容和次重点内容均需理解记忆，另外一些不需记忆但是需要理解的，比如主题标引的实践，文档并未列出，参见 ppt。

复习时建议把书也看一遍，按照重要程度为①第四章主题法、第五章分类主题一体化词表②第一章导言、第二章检索语言原理、第六章文献标引概述③第八章主题标引和主题检索工具、第九章自然语言标引与检索。

FionaY

2020 年 1 月

目录

写在前面.....	1
第一章 情报语言学概述.....	3
一、引言.....	3
二、情报检索语言概述.....	3
三、情报检索语言的基础.....	4
四、回顾 IRR (Information Representation and Retrieval)	6
第二章 情报检索语言原理.....	7
一、情报检索中的词汇控制.....	7
二、情报检索语言中的句法控制.....	8
三、检索效率和词汇控制.....	9
第三章 主题语言.....	11
一、主题语言概述.....	11
二、叙词法的词汇控制.....	13
三、叙词表的结构.....	15
四、叙词表的使用——如何查词?	18
五、叙词表的编制与管理.....	19
六、叙词表的评价.....	22
七、国内外主要主题词表简介.....	23
第四章 分类主题一体化语言.....	25
一、分类主题一体化词表.....	25
二、分面叙词表.....	27
三、分类主题词表.....	28
第五章 主题标引.....	32
一、主题标引概述.....	32
二、主题标引的流程.....	33
三、主题分析.....	33
四、主题标引的一般规则.....	34
17 级题型.....	38
一、名词辨析 4*4 个.....	38
二、简答题 6*9 个.....	38
三、论述题 10*3 个.....	39
17 级考试题目.....	39
一、名词辨析 4'	39
二、简答题 6'	39
三、论述题 10'	39

第一章 情报语言学概述

一、引言

二、情报检索语言概述

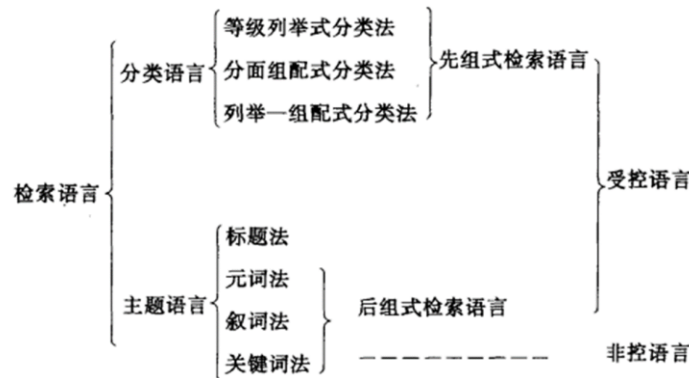
※1. 情报检索语言的定义、性质和分类

(1) 情报检索语言的定义

情报检索语言 (Indexing Language)，又称检索语言或标引语言，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，是根据信息检索的需要而创制的一门人工语言。

书：是一种由表达文献主题概念及其相互关系的词汇及规则组成的人工语言系统。

(2) 情报检索语言的分类



①人工语言与自然语言

人工语言：人类根据自定的一套规则编写而成的，其规则是预先明确规定的。

自然语言：人类在交际中自然演变而成的，其规则反应预先被明确规定的社会流行用法，是约定俗成的。

书：

受控语言：依据特定词表或类表揭示文献信息的组织方法。特点：标引和检索均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结果。

自然语言（非控语言）：直接使用文献或用户检索使用的自然语言进行组织的方法。特点：不需要使用受控词表，但一般仍需要遵守一定的文献标引规则或检索措施，以改进使用效果。

②分类语言与主题语言

分类语言：以学科性质为主对各种概念加以分类和系统排列，并用一定的标记符号加以表达的文献信息检索语言。《中国图书馆分类法》

主题语言：采用描述文献主题内容的词语标识并按字顺排检文献的检索语言。《汉语主题词表》

书：

分类法：以分类号码为检索标识，按照文献内容之间的相互关系加以组织的信息检索系统。包括等级列举式分类法、分面组配式分类法、列举-组配式分类法等。

主题法：直接以语词表达信息内容，以字顺或其他方式加以组织的信息检索系统。包括标题法、元词法、叙词法、关键词法等。

（3）情报检索语言的性质

检索语言=词汇+语法

词汇：登录在分类表、词表、代码表中的全部标识（分类号、检索词、代码）。

语法：如何创造和运用这些标识来正确表达文献内容和情报需要，以有效实现情报检索的一套规则。

书：

词汇集是检索语言的主体，一般用于对所使用的词汇及其相关性进行规定，是文献存储和检索的语言保障。其表现形式是各种词表，包括分类表或标题表、叙词表等。

语法是分类标引和主题标引的规则系统，它们规定词汇集的使用方法，是文献标引必须遵守的规范。

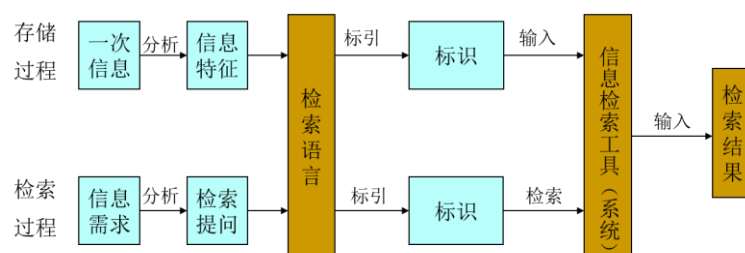
※2. 情报检索语言的优势

- ①简单明白又比较专指地表达文献及检索课题的主题概念
- ②语词与概念一一对应，排除了多词一义、一词多义和词义模糊现象
- ③显示概念之间的关系
- ④将概念进行系统排列
- ⑤检索时便于将标引用语和检索用语进行相符性比较

如：同义词环 Synonym Rings、等级分类体系

同义词环相对来说是一种较新的结构；通常与检索引擎一起使用，帮助扩充检索式；用户只要使用了同义词环中的任何一个词，就可以找到这一类的资源。

※3. 情报检索语言的作用



4. 情报检索语言的功能（知识组织）

- ①对信息内容加以标引的功能；
- ②对内容相同及相关的信息加以集中或揭示其相关性的功能
- ③对信息资源加以系统化或组织化的功能
- ④便于将标引用语和检索用语进行相符性比较的功能

三、情报检索语言的基础

☆情报检索语言的基础：术语学、概念逻辑、知识分类

1. 术语学

(1) 术语学

术语学是一门综合了信息科学和语言学，专门研究概念的订名、应用及概念间相互关系的交叉学科。术语学是指导术语标准化的重要工具。

(2) 术语学与情报检索语言

- ①一种检索语言就算一个经过精心组织的术语集
- ②术语是分类表、词表的要素
- ③应用与被应用的关系
- ④检索语言以术语学对科学术语的研究成果为依据

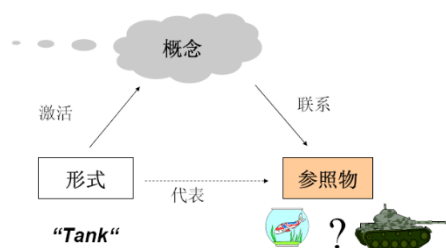
2. 概念逻辑

(1) 概念逻辑

- ①概念是事物本质属性的概括。
- ②概念逻辑是一种科学思维方法，揭示事物的本质属性及各种事物之间的区别和联系。
- ③任何一种检索语言都是建立在概念逻辑基础上的，通过各种检索语言的语词（标识）来表达概念。
- ④是否符合概念逻辑是衡量检索语言质量的重要标志

★ (2) 概念与语词

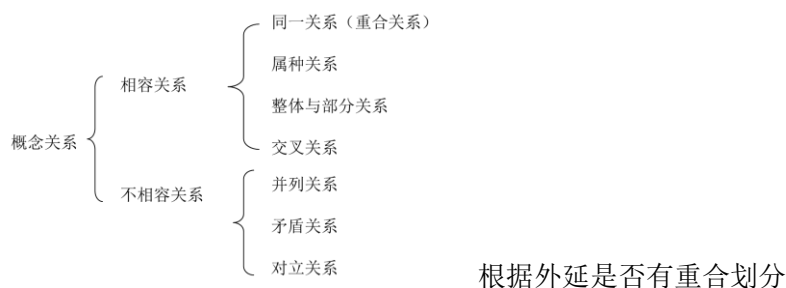
- ①概念和语词的联系：语词是概念的语言形式，概念是语词的思想内容。任何概念都是通过语词来表达，任何和意义的语词所表达的含义都是概念。
- ②概念和语词的区别：概念属于思维科学的研究对象；语词属于语言学科的研究对象。语词并不都表达概念，比如，如果、所以等语词就不表达概念。对应关系不同，同一概念可以用不同的语词表达，如医生、郎中表达的是同一概念；同一语词可以表达不同概念，比如白头翁，即可以表达一种鸟，也可以表达一个老头。



(3) 概念的内涵和外延

- ①内涵：又称为含义，就是反映在概念中的对象的本质属性，即概念的质的规定性。其作用是表明对象“是什么？”。在日常用语中通常用……是……、……即……、……就是……等语句表示概念的内涵。
- ②外延：指具有概念所反映的本质属性的全部对象，即概念的量的规定性。其作用表明对象“有哪些？”。在日常用语中通常用……包括……、……有……、……可分为……等于局表示概念的外延。

(4) 概念之间的关系



(5) 检索语言应用的两种概念逻辑方法

①概念的划分与概括（分类）

概念的内涵和外延成反比关系，用来区别客观事物，在划分中产生了隶属和并列关系，最终形成——概念等级体系

②概念的分析与综合（组配）

概念交叉关系中，外延重合的两个概念可以形成一个新的概念，与原概念具有隶属关系。或者两个外延较浅的概念综合为一个内涵较深的概念，最终形成——概念组配体系。

3. 知识分类

①知识分类是根据知识的内在联系进行科学组织，以便系统地进行检索、利用的主要方法。

②检索语言中的体系分类法、组配分类法、叙词法以及标题法等都在不同程度上反映了知识分类。

③知识分类的多维性：早期检索语言、超文本

四、回顾 IRR（Information Representation and Retrieval）

✖1. 信息表示和信息检索（IRR）的发展脉络

①Increased Demand (1940s-early 1950s)

二战期间产生的大量科技文献

②Rapid Growth (1950s-1980s)

词间关系控制（专业人员检索）

③Demystified Phrase (1980s-1990s)

去神秘化，终端用户发生变化

④The NetWorked Era (1990s-present)

形成自然语言与受控语言的集成化：用户自然语言检索，系统受控语言标引，中间情报检索语言控制

★2. 补充：用户和信息环境如何对情报语言产生影响

①受控语言产生之前：自然语言是唯一的 IRR 语言，人们开始意识到自然语言的种种弊端。

②受控语言的产生：先组式到后组式；自然语言和受控语言同时存在，开始产生争论。

③关键词和全文检索时期：自然语言再度热起来。

④集成化时代：自然语言应用于前台，后台受控语言；终端用户不再接触受控词表；集成化，各司其职。

印刷时代，两种语言并行；联机时代，更侧重使用受控语言；数字时代，自然语言大行其道

第二章 情报检索语言原理

一、情报检索中的词汇控制

※ (1) 词汇控制定义

词汇控制：指根据文献标引和检索的需要，对自然语言中的词汇进行选择、规范并揭示其相关性的过程。

(2) 自然语言的特点

- ①使用灵活、表达性强；
- ②含义不够精确，缺乏明确的结构，词汇量过大；
- ③语词与概念不是一一对应的，存在大量一词多义、一义多词及词义含糊的现象；

(3) 检索语言的词汇要求

单义性、准确性、适用性、相关性、兼容性

※ (4) 词汇控制的范围

词量控制、词类控制、词形控制、词义控制、词间关系控制、专指度控制、先组度控制 P19

1. 词量控制

常用的词量控制方法：

- ①以一个宽泛的上位词，代替若干专指的下位概念
- ②通过表示单元概念语词的组配，表达复合概念来代替直接收入词组，减少词量。

2. 词类控制

- ①控制、排除与标引和检索无关的词汇。
- ②按照对文献主题内容揭示的要求，一般只使用名词或名词性词组。必要时使用形容词、介词、连词和数量词。

3. 词形控制

一个概念只用一个语词表达，包括同义词、准同义词、词形的选择和规范等。
汉字形体的控制、外来词的控制、标点和数字的控制、同义词的控制、

4. 词义控制

加限定词、加注释

5. 词间关系控制

	分类法	主题法
词间关系	等级结构、类目注释、类目参照、交替类目	参照系统
等同关系	类名后的限定词或类目的同义词注释	用 use、代 UF
等级关系	类目的等级体系、版面的字体、字号、缩格等辅助手段	属 BT、分 NT
相关关系	参见注释和内容范围注释	参 RT

6. 专指度控制

★专指度：检索词与文献主题概念的切合程度，是影响检索语言标引质量的一个重要因素。
较高专指：有助于提高检索的准确度，满足对特定主题文献的查找。
过度专指：造成词量过大，增加标引和检索的难度，造成文献分散，影响检全率，增加系统的管理费用。

（1）分类法中的专指度

类目划分层次决定着分类体系的深细程度、类目的专指度。
划分层次越多，聚类的精确性越高，按某种属性检索的精度就越高。
类目专指度高，知识体系越复杂，相关的知识越分散，一些深细的知识隐藏不易被检索到。
“集中与分散”的矛盾

★（2）如何提高/降低专指度？

提高专指度的措施	降低专指度的方法
①扩大专指词的选用； ②扩大收词范围，改进收词的完备性； ③增加类表或标题的级数或组配的级别； ④允许采用自由词补充标引； ⑤提高词汇的先组度等。	①合并近义词，用含义接近的词代替； ②采用上位词置代，以比较宽泛的词标引过度专指的概念； ③减少类表或标题级数或组配级别等； ④以组配形式代替先组词等。

7. 先组度控制

★先组度：检索词汇的先组程度，包括采用先组方式，还是后组方式，以及词表中先组词的使用情况。
词组的数量决定先组程度。

（1）分类法中的先组度

体系分类法属于先组定组式，分类号是一个不可拆分的整体，难以进行组配检索。
分面分类法依据概念的分析与综合原理，将主题概念组成“分面—亚面—类目”的结构体系，通过各分面内类目之间的组配来表达文献主题。

★（2）补充：提高专指度/先组度对信息检索/标引的影响

改善检索系统对文献主题的标引精度，提高检准率；增加标引的难度，降低标引的一致度，影响检索系统的检全率。
使用专指词过多，会增加系统词量以及相关专指概念之间的区分难度，影响标引的准确性，降低系统的检准率。专指标引使一些总论性或相关文献被排除在外，从而影响检索系统的检全率。

二、情报检索语言中的句法控制

语词关系=语义关系+句法关系

1. 句法控制概述

句法控制：根据文献标引和检索的需要，通过一定的组词造句规则，对检索语词的组合方式作出规定。简单来说，就是检索语句中语词之间形成的关系。
自然语言中，由词序和虚词来表示句法结构。

✖检索语言中，采用引用次序、辅助符号、轮排等基本形式表示句法关系。

2. 引用次序

✖引用次序（组配次序）：对复合主题进行标引和检索时，各个主题因素的组合或排列的次序。

在分类法中，是指类目划分标准或不同组面被引用的先后顺序；引用次序决定着类目体系集中文献的方式以及检索途径，体现了分类法的性能和适用性。

在主题法中，是指字顺标题中各个主题因素的排列次序。是先组式主题语言进行句法控制的重要手段。

（1）显著性引用次序

根据主题概念对用户心理映象的显著程序确定的次序。

事物——部件——材料——活动——施动者

（2）基本范畴引用次序

基本范畴引用次序是指将各种主题概念划分为范畴，并按范畴的性质、职能确定组配顺序。如：印度图书馆学家阮冈纳赞的范畴分面公式、英国分类法研究小组维克利提出的标准引用次序。

阮冈纳赞概括为：本体、物质、动力、空间、时间五大基本范畴。确定了 PMEST 这一引用公式。

（3）上下文从属引用次序

上下文从属引用次序是一种接近自然语言句法手段和语序的引用次序。这一引用次序是随着索引编制自动化的发展逐步建立起来的。

✖（4）中文引用次序

刘湘生将主题因素归纳为五个方面：主体因素（A）、通用因素（B）、位置因素（C）、时间因素（D）、文献类型因素（E）。

★补充：引用次序对检索结果的影响

通过明确主题概念之间的句法关系，排除不符合检索要求的组配关系，提高检索的准确度；增加标引和检索的难度，提高系统的费用。

三、检索效率和词汇控制

检索效率表示对检索系统各项性能的满意程度。

检索效率涉及到多种因素，包括：检全率、检准率、响应时间、检索费用等各种要求。

检全率、检准率是评价检索效率的两个最基本的指标。

1. 检全率

检全率：亦称查全率、命中率，是指通过检索系统检出的与某一检索提问相关的文献数与检

索系统中与该提问相关的实有文献总数之比，可用下式表示：
$$\text{检全率} = \frac{\text{检出相关文献量}}{\text{系统相关文献总量}} \times 100\%$$

2. 检准率

检准率：亦称查准率、相关率，指通过检索系统检出的与某一检索提问相关的文献数与检出文献总数之比，可用下式表示：

$$\text{检准率} = \frac{\text{检出相关文献数}}{\text{检出文献总数}} \times 100\%$$

$$\text{查全率: } R = \frac{\text{被检出相关文献量}}{\text{相关文献总量}}$$

$$\text{查准率: } P = \frac{\text{被检出相关文献量}}{\text{被检出文献总量}}$$

$$\text{漏检率: } M = \frac{\text{未检出相关文献量}}{\text{相关文献总量}}$$

$$\text{误检率: } N = \frac{\text{被检出非相关文献量}}{\text{被检出文献总量}}$$

※3. 词汇控制措施对检索效率的影响

词汇控制措施	提高检全率	提高检准率
词类限定	√	
词量控制	√	
提高先组度		√
提高专指度		√
词形规范	√	
同义词优选	√	
准同义词控制	√	
等级关系控制	√	√
相关关系控制	√	√
引用次序		√
控制符号		√
轮排		√

书：

①多数检全措施具有降低标引和检索的专指度的特点。有助于降低标引和检索的难度，将涉及该专指主题的总论性文献或相关文献检出，但检全一个主题概念的文献，必然会在同时影响检准率。

②词间关系控制：通过揭示语词之间的联系，提供选择检索的能力，可以改善检索系统的灵活性，为用户提供了通过词间关系进行扩检、缩检、改变检索方向的能力，提高了检索系统的整体效能。标引者和用户可以通过系统揭示的联系，逐步缩小其检索范围，提高检准率；也可以根据实际检索情况，利用词表对相关概念的揭示，扩大相关文献的查找，提高检全率。

第三章 主题语言

一、主题语言概述

★1. 基本概念

主题：文献论述的对象，包括事物、问题、现象等。

主题词：经过选择，用来表达文献主题的语词。

主题法：直接以表达文献主题的语词作为检索标识，以字顺为主要检索途径，并通过参照系统等方法揭示词间关系的标引和检索文献的方法。

2. 主题法的特点

- ①以主题为中心集中文献（分类法是以学科体系为中心集中文献）
- ②直接以语词作为检索标识（分类法以号码系统为检索标识）
- ③以字顺方式作为主要检索途径（分类法以类目之间的关系作为检索途径）
- ④通过详尽的参照系统揭示主题词之间的关系
- ⑤用来处理文献，编制各种检索工具及检索系统（分类法用来组织文献排架及编制分类检索工具）

3. 主题法的类型（发展脉络）

✖按照构成原理分：标题法、元词法、叙词法、关键词法

✖按照标识的组配特点分：先组式主题法、后组式主题法。

（1）标题法

标题：经过规范化处理的，用来标引文献的词或词组，通常为比较定型的事物名称。

标题法：以标题词作为文献主题标识的标引和检索的主题法。

标题法是最早出现的一种主题法，属于**列举式**主题法，先组式

1) 标题的类型

单级标题、词组标题、多级标题、带限定词的标题、倒置标题

2) 标题法的词间关系

通过参照揭示**词间关系**，被叙词法所吸收

在 LCSH 中：use（用）、UF（代）、BT（属）、RT（分）、RT（参）

✖3) 标题法的特点

- ①列举式词表，形式直观；
- ②定组式标题，结构固定，含义明确；
- ③按词表列举的标题标引，操作简便
- ④收词量大；
- ⑤专指度不足，修订量大；
- ⑥只能从规定的组配顺序入手查找，无法从多因素、多角度检索。

（2）元词法

元词法：一种以元词作为主题标识，通过**字面组配**的方式表达文献主题的主题法。

元词：用来标引文献主题的、最基本的、字面上不能再分的语词。

元词法是后组式语言的先驱，是机械检索系统中**倒排档**的先驱。

✖1) 元词法的特点

- ①词表体积小
- ②标引专指度高
- ③便于从不同角度检索
- ④字面组配法，在字面和语义不一致时，容易造成误差；
- ⑤缺乏完善的参照系统，不利于族性检索。

（3）叙词法

叙词法：从自然语言中精选出来的，经过严格处理的语词作为文献主题标识，通过概念组配表达文献主题的主题法。

吸收了元词法、标题法、分面组配法的优点。形成于 50 年代末，60 年代初期。

叙词（Descriptor）：又称主题词，是经过规范化处理的，以基本概念为基础的表达文献主题的词和词组。

1) 叙词法的发展简史

※2) 叙词语言对其他检索语言的吸收

表 4—1 叙词语言对其它检索语言技术的吸收

其它检索语言	技 术	叙词语言技术
体系分类语言	学科分类 等级结构	范畴索引 词族索引
分面分类语言	分面组配	概念组配
标题语言	先组 见参照、参见参照	适当先组 用、代、属、分、参参照
元词语言	字面组配 完全后组 反记法	概念组配 基本后组 倒排档
关键词语言	轮排技术	轮排索引

※3) 叙词语言的特点

- ①组配准确，标引能力强；
- ②结构完备，词汇控制严格；
- ③适合多途径检索，检索效率高；
- ④对检索系统的适应性强。
- ⑤词汇控制严格，词表编制难度大；
- ⑥标引规则复杂，标引难度大，速度慢；
- ⑦用户难以熟悉词表及标引规则，给使用带来不便。

（4）关键词语言

关键词语言：以关键词作为文献内容标识和检索依据的一种信息检索语言。

※关键词：从文献题名、文摘或正文中抽取的，能够表达文献主题并具有实质意义的未经规范化处理的自然语言词汇。

关键词表达事物概念，具有直接、准确等特点。

背景：标题法、元词法和叙词法都是受控语言（人工语言），在互联网时代不被用户接受，因此关键词法（自然语言）大行其道。这时的检索是基于关键词匹配的，只要输入的关键词在目标文献中出现的频率高，就视为正确的检索结果；存在歧义，单纯的关键词语言无法实现检索目标。因此前端-自然语言；后端-受控思想。

（5）补充：信息检索和信息组织的三个网络时代

Web1.0 时代：文件网

特点：互联网发布信息，用户查看。关键词法。

Web2.0 时代：人际网

特点：信息发送方和接收方对等，协同众包。

Web3.0 时代：数据网

特点：提出 semantic web

4. 叙词法的原理——概念组配

★任何一种在自然语言基础上发展起来的主题法都会遇到一个不可克服的矛盾，那就是收词量有限的词表与海量的自然语言词集之间的矛盾。因此，必须通过精心选择和采用语词组配这一重要的原理和方法来表达各种文献的主题概念。



※ (1) 字面组配 - 拆词

字面组配：根据构词法的原理，将一个复杂的、专指的概念按其字面构成拆分成两个或两个以上的单词，标引的时候在根据构词法的原理将两个词拼接起来，表达一个专指、复杂的概念。

※ (2) 概念组配 - 拆义

概念组配：根据概念的分析与综合原理，将一个专指、复杂的概念按其语义分解为两个或两个以上的较为宽泛的概念，标引时同样按其语义将两个或两个以上的较为宽泛的概念组合表达一个专指、复杂的概念。

★ 2) 概念组配的类型

- ①交叉组配：即内涵不同、外延部分重合的两个相同性质概念之间的组配。组配的结果产生一个新概念，这个概念分别属于这两个概念的下位概念。
- ②限定组配：即两个不同性质概念之间的组配，其中一个概念反映了另一概念的某一方面、某一特征或时空中的某一部分。限定的结果也产生一个新概念，它表示该事物的某一方面或某一特征。
- ③联结组配：这种组配表示几个概念之间的联系，并不形成新的概念，只揭示参加组配概念之间的某种联系。

※ 3) 叙词组配的作用

- ①编制叙词表时，可以控制词量、缩小词表篇幅。
- ②标引时，通过组配来保证标识表达主题概念的专指性。
- ③提供多途径检索。
- ④有助于自由扩大或缩小检索范围。
- ⑤能比较及时反映新事物、新学科。

二、叙词法的词汇控制

1. 词类控制

- ①词类控制：以简单名词和名词性词组为主，收录少量形容词。
- ②词组控制：词组分解后用于组配会产生歧义的或不合用户习惯的则选择词组作为叙词。

如：专有名词性词组、专业文献中出现频次较高的、分解后没有单独意义的词组

2. 词形控制

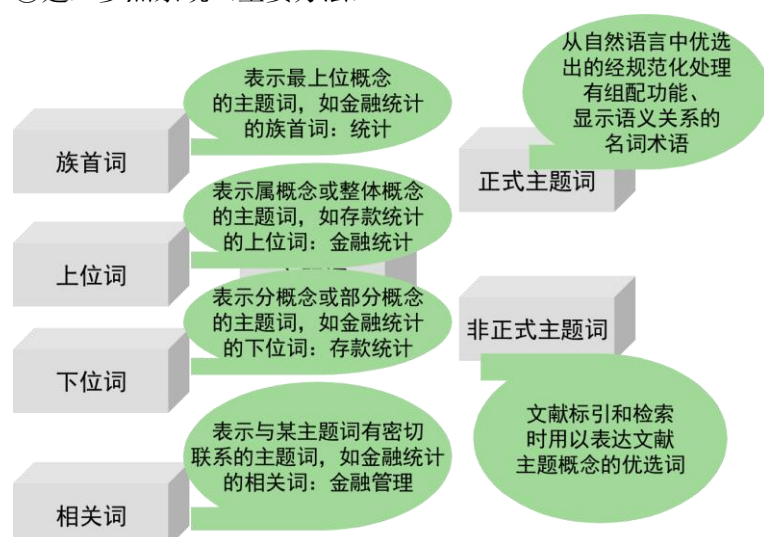
- ①汉字形体的控制：选择通行的或权威规定的写法。
- ②外来词的控制：一般采用通用的汉语译名，除非是公认的，比译名更流行。
- ③标点和数字的控制：圆括号、连字符和小圆点，其他符号一般不用
- ④同义词、准同义词的控制：选一个作正式词，其余为非正式词。

3. 词义控制

- ①加限定词
- ②加注释：含义注释、用法注释、历史注释

4. 词间关系控制

- ①编制分类性质的索引（如范畴索引、词族索引）；
- ②运用图示方法（如词族图等）；
- ③建立参照系统（主要方法）。



(1) 等同关系

等同关系：正式叙词和非正式叙词之间的关系。

用 (Y)：非正式叙词指示符，非正式叙词指向正式叙词。

代 (D)：正式叙词指示符，正式叙词指向正式叙词。

作用：集中同一主题文献，增加检索入口

使用范围和方法：在同义规范的基础上进行，一般应用参照和反参照

(2) 等级关系

等级关系：专指度不同的两个叙词上位叙词和下位叙词之间的关系。

属 (S)：下位叙词指示符。

分 (F)：上位叙词指示符。

族 (Z)：族首词，族系关系。

功能：扩大或缩小检索范围

使用范围：属种关系、整部关系、集合概念与个体概念

整部关系：揭示地理位置中的整部关系、揭示人体系统与器官以及某些事物中的整部关系、

揭示组织机构及分支中的整部关系

（3）相关关系

相关关系：又称类缘关系，在概念上属于相互关联、交替、矛盾、对立等地位。

对标引人员选词很重要，同时对提高检索效率也有很高的参考价值。

参（C）：相关关系

基本类型：近义关系、反对或矛盾关系、交叉关系、事物与其部分/成分、因果关系、应用关系、学科或研究领域与其对象、影响关系、各种事物与材料/性质、理论/学派/组织/运动/事件与有关的人物或团体、组代词

※（4）词间关系符号

词间关系	符号含义	汉语拼音符号	英文符号
等同关系	用	Y	Use
	代	D	UF
	组代		UFC
等级关系	分	F	NT
	属	S	BT
	族	Z	TT
	（属种）属		BTG
	（属种）分		NTG
	（整部）整		BTP
	（整部）部		NTP
等级关系	参	C	RT

（5）先组度控制

先组度控制：对词汇先组程度的控制，是叙词表编制中的难点，组配过多或过少都影响词表性能。

（6）利用叙词表的词间关系检索

扩检（Broader Term Search，即广义词检索）

缩检（Narrower Term Search，即狭义词检索）

联检（Related Term Search，即关系词检索）

族检（Term Family Search）

群检（Narrower Block Search）

（7）词间关系可视化

Protégé 是由斯坦福大学医学院的医学情报学研究组(Stanford Medical Informatics)开发研制的一个免费的开源平台。

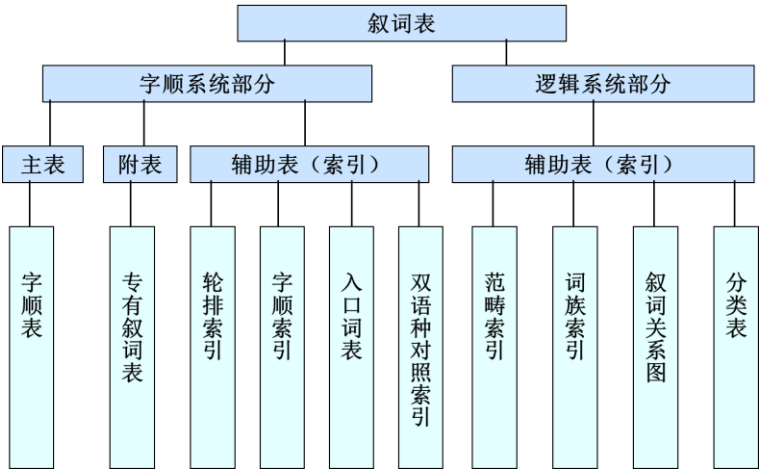
Protégé 具有一套丰富的知识构建框架和方法，能以各种表示格式创建图形显示和管理本体。

三、叙词表的结构

叙词表（Thesaurus），又称为主题词表，它将自然语言转换为规范化的系统语言的一种术语控制工具，概括由自然语言优选出的语义相关、族性相关的学科术语所组成的一种规范化动态词典。

✖1. 叙词表的宏观结构

按其显示方式，叙词表有：字顺显示、系统显示、图形显示三种方式。
按照各组成部分的性质，一个完整的叙词表一般包括字顺表、附表、辅助索引等。



✖（1）字顺表

字顺表：由叙词款目和非叙词款目按字顺排列的一览表。是叙词表的主表。
功能：供标引员和用户按照字顺直接对叙词进行查找，并通过参照系统判定词汇含义，进行相关词的查检。是标引和检索的主要依据。
收词以普通叙词为主，也选择的收录一定数量的常用专有叙词——专有叙词表

（2）附表——专有叙词表

是将众多专有叙词款目按字顺排列的一览表，是字顺表的重要组成部分。
专有名词指表达某一单一事物对象的名称。常用的专有名词包括人名、地名、机构名、产品型号名等，数量大，范围广
作用：①缩小主表篇幅；②分类显示作用；③有利于管理或增补。
款目结构和排列方式与字顺主表相近。一般不收入范畴索引，无范畴号。
只有在必要时才编制。

（2）附表——人物表

（2）附表——组织机构表

（3）双语种对照索引

是一种将全部叙词、非叙词与相应外语译名相互对应，按外文译名字顺排列的词汇表，是通过外文译名使用词表的辅助工具。
按外文字顺排列
便于标引人员和检索人员从外文反查汉语叙词
只列出“代项”和“参项”

✖（4）范畴索引

是一种按照词汇所属学科或专业范畴编制的概略分类系统，是从分类角度查找叙词的辅助工具，亦称分类索引。一般涉及如下内容：
①索引中类目的设置，一般以学科分类为基础，结合专业分类和词汇分类特点进行；

- ②范畴类目的组织，一般有：按类目层次和内容关系组织，类似等级分类体系；按分类字顺方式组织，即按字顺排列类目体系；
- ③类目标记代码，通常采用少数字母和数字混合的方式，一般需要考虑代码的扩充性；
- ④叙词和非叙词的收录，一般收录字顺表中的全部叙词和非叙词，所有词均按照分类体系进入索引中的相应类目，原则上一词一类。

※ (5) 词族索引

词族，指一组具有属分关系的叙词结合。词族索引是一种以词族为款目单元，按照款目词的字顺排列，可以从等级关系的角度查找叙词的索引。一般的处理包括：

- ①词族索引只显示词表中按属分关系处理的叙词；
- ②族首词的选择，一般选择外延较宽，具有实际检索意义的叙词作族首词；
- ③词族的规模，一般处理方法：对词族的规模不进行限制，对等级关系进行完整显示；将规模过大的词族分设几个词族，总词族和分词族均作为独立的词族分别排列；
- ④词族的显示，一般以族首词为标目，按等级关系展开；
- ⑤当一叙词同时从属于两个或多个词族的处理，一般在相应词族下同时予以反映，以便使用者能从不同角度查找该词。

(6) 轮排索引

将词表中的叙词按词素的字顺排列，使含有同一词素的叙词集中显示，提供从词素的角度查找叙词。一般的处理内容：

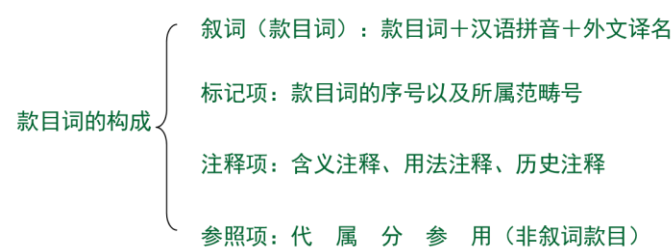
- ①收词范围，一般选择具有词组形式的叙词和非叙词，也可将单词同时收入轮排索引；
- ②每个叙词按照构成的词素，确定轮排条目数量；
- ③按照标目的显示方式，有词内轮排、词外轮排、倒置轮排三种轮排形式；
- ④排序方法，通常有顺序排列和逆序排列两种排列方法。

(7) 词族图

将叙词之间的关系用二维图的形式加以显示，使用户可以通过直观的图像，迅速了解词间关系。一般有：树型显示图、箭头显示图、环形显示图、其他。

※ 2. 叙词表的微观结构

叙词表的微观结构是指叙词款目的构成。叙词款目是叙词表的基本单元。



(1) 叙词款目

叙词——	Nongye zhengce	——汉语拼音
	农业政策	05B——范畴号
	Agricultural policy	——英文译名
代项——	D 农业方针政策	——非正式叙词
分项——	F 副业政策	
	粮食政策	—— 下位词
	林业政策	
	...	
属项——	S 经济政策	——上位词
族项——	Z 政策	——族首词
参项——	C 农村经济政策	
	农业法令	—— 相关词
	...	

(2) 非叙词款目

Nongye fangzhen zhengce
农业方针政策 05A
Agricultural program and policy
Y 农业政策

3. 叙词表的组成模式

一个功能完善的叙词表至少包括两部分：字顺表+范畴表（或分类表）

（1）好词表应具备的功能

- ①功能完备。各个组成部分的组合合理。
- ②各部分功能应相互补充。
- ③结构简明。各组成部分避免重复。
- ④编制和管理方便。
- ⑤易于掌握、便于使用。

★四、叙词表的使用——如何查词？

1. 字顺表的使用

适用于主题概念比较稳定，词形已经确定的情况。

必要时再通过词族索引进一步查核。

2. 轮排索引的使用

适用于词形一时无法确定的主题概念，但是知道其包含某一词素。

应再通过主表查核。

3. 范畴索引的使用

某些文献的主题概念集中于一个学科；学科归属明确，但语词形式一时无法确定或词表中没有对应的词，需选用近义词标引时。

4. 词族索引的使用

用于从等级角度查找对应的专指叙词。弥补字顺主表中对等级揭示不完备的缺憾。

5. 英汉索引的使用

用于标引英文文献时。

五、叙词表的编制与管理

※1. 叙词表的编制程序

总体合计

制定编制规范

拟定类表收集词汇

整理词汇

建立词间关系

编表

审核和试标引

定稿出版

(1) 总体设计

1) 需求分析

①调查现有词表满足需要的程度

②明确拟编词表的用途:

- 使用对象: 某特定机构、系统、数据库、社会上各有关机构?
- 使用环境: 手工检索、计算机检索?
- 使用目的: 分类、标引还是二者兼而有之?
- 词表类型: 受控、自然语言、二者结合?

2) 明确学科范围, 明确标引对象

- 学科: 重点学科、相关学科、边缘学科?
- 标引对象: 图书、期刊论文、科学报告、档案、网络资源?

3) 确定词汇特点

确定收词的专制度和先组度

4) 规定词表结构

①确定宏观和微观结构。

②宏观: 主表的显示方式, 设置哪些附表、辅助表的种类?

③微观: 主表、附表的款目结构及构成元素

5) 明确编表条件

编表所需的人力、财力、设备和时间, 合理安排资源

(2) 制定编制规范

①对编表中涉及的理论和技术问题作出具体规定

②参考叙词表编制的国家标准和国际标准

③选词的原则和范围

④词汇控制的内容和原则

⑤参照项目的设立, 参照项目符号含义

⑥词汇数量及词汇先组度和专指度

- ⑦主表、附表、辅助表的著录项目、格式和排列规则
- ⑧注释的类型及撰写方法

(3) 拟定类表收集词汇

1) 叙词表的收词原则

- ①目标性原则
- ②文献保证原则*
- ③用户保证原则*
- ④组配优先原则
- ⑤基本词汇完备与精练原则
- ⑥兼容性原则

※2) 词汇收集

按词表涉及的学科和专业分类列表收集，起到分工协调、查检遗漏作用；为编制分类索引打基础。

①经验法（empirical approach）自底向上

收集该学科的百科全书、字（辞）典、刊物及其他权威性出版品，将出现的词记录于作业表格，而后再进行属性划分。亦称“石笋式法”（stalagmitic）

词汇收集来源：现有的词表（标题表、叙词表）和分类表；检索工具，尤其是主题索引；词库；工具书（专业词典、百科全书、术语手册等）；原始文献（专著、论文、技术报告、标准文献等）

②小组汇编法（committee approach）自顶向下

汇编法则是聚集一群学科专家，进行头脑风暴，列出相关词汇，确定族首词（top term），再行分族编排。又称为“钟乳石法”（stalactitic）

(4) 整理词汇

- ①对词汇进行分类，按照学科、专业集中
- ②同一学科、专业类目下，按照基本范畴或分面分类
- ③对各范畴或分面下的词汇进行比较分析，剔除重复，确定词间关系，形成叙词表的基础

★(5) 建立词间关系

1) 等同关系

又称用代关系，参照符号是「USE」及「UF」

同义（synonyms）指词义完全相同，可互相取代的词

准同义（near-synonyms）指意义相近

意义不同但为了标引目的视为同义的词

组代关系 标引时以数词组合代表另一概念

描述词 V.S.非描述词

范围注释与解说

2) 等级关系

又称属分、上下或层级关系，参照符号是「BT」及「NT」

类别：属种（genus-species）、整部（whole-part）、多层次（polyhierarchical）

判定公式

属种关系：“一些（鸟类）是（鹦鹉）”，“所有（鹦鹉）全都是（鸟类）”

整部关系：“（江苏）的一部分是（南京）”，“（南京）是（江苏）的一部分”

3) 相关关系

又称亲缘、类缘或联想关系，参照符号是 RT

判定参考原则

同一范畴：属性含义有部分重叠但非同义的词可互为 RT

不同范畴：两次虽分属不同的概念体系，但彼此间却有强烈的提示关系，如学科及对象（林学和森林）、事件与相关人物等等

（6）编表

生成完整的词表，包括字顺表和各种索引

要求排列正确，相关参照完整，等级显示明确

（7）审核和试标引

1) 审核

审核要对词表的所有步骤和元素进行审查

自查或专家审查

2) 试标引

对保证词表质量非常重要

待标引的测试文献要有足够的数量以及涵盖词表的各个主题

试标引后，可以对词表进行进一步修改、补充和完善

（8）定稿出版

要有编制说明和使用方法说明

2. 叙词表的管理维护

实际上就是对词表进行修订。增补新词、剔除旧词、修改调整。

3. 计算机在词表编制和管理中的作用

（1）优越性

- ①便于原始数据的收集和整理，保证编表的质量与速度；
- ②可自动生成各种结构和显示方式，节省人力和物力，降低编制成本；
- ③可快速生成各种版本和各种载体的词表，方便用户选择；
- ④便于词表的管理维护，增强词表的使用功能和效果；
- ⑤为词表与联机系统和网络检索的连接提供可能

（2）计算机在词表编制中的应用

1) 词汇收集

- ①直接从数据库中获取标引词、检索词及使用频率，作为词汇选择的重要依据
- ②直接从文献题名、文摘、正文中抽词，供词表选词参考
- ③记录人工收集的词汇，进行统计，分析，为词汇优选，词间关系的确定提供参考

2) 词表辅助生成

- ①自动生成款目，反参照
- ②自动生成词表的各种显示
- ③排版输出，为词表生成印刷数据

（3）计算机在词表管理和维护中的应用

- ①可随时进行叙词的增、删、改，为词表的增、删、改积累数据
- ②在增、删、改中确保词间关系和词序的正确

(4) 计算机主文档的应用

- ①提供联机显示，便于用户查词选词
- ②统计叙词的标引频率和检索频率，为检索策略优化和词表管理提供依据
- ③将入口词自动转化为受控词，进行标引和检索，并在系统需要时，进行上位登录
- ④依据词表中的参照和等级结构，自动扩检和缩检

4. 词表编制软件——Multites 使用

六、叙词表的评价

1. 叙词表评价的意义

(1) 张琪玉教授认为，决定检索系统检索效率的因素：检索语言的质量、标引质量、检索质量及其他因素。

(2) 评价的意义：

- ①可以发现词表在标引和检索中存在的问题，为词表修订提供依据；
- ②考察词表能否满足检索效率对检索语言提出的要求，为词表现代化提供参考
- ③便于从定性和定量角度全面了解词表，为用户选择合适的词表提供参考依据

※2. 叙词表评价的内容

(1) 宏观评价

主要是对词表结构的评价

字顺表

词表各组成部分相互对应、相互参照的程度

词表结构设计和实现的难易程度

(2) 微观评价

主要是对词表词汇的评价

词汇的数量

入口词的比率

先组度

参照度

关联比

3. 叙词表评价的方法

(1) 定性评价

完整性：结构的完备（系统显示+字顺显示）

准确性：术语的科学性

动态性：更新的速度和容纳性

排印质量和产品多样化和系列化：版面以及不同类型版本

管理手段和体制的完备性：专门的管理机构、管理系统

✖ (2) 定量评价

1) 词量

词表包括词汇的多少。

是衡量叙词表词汇完备性和专指度的指标，也是对叙词表规模划分的依据。

叙词表收词丰富，专指度深，能充分标引文献。

词量受词表包含的学科范围、文献数量、先组度等因素制约。

2) 等同率

指非叙词与叙词的比率。这是衡量入口词汇丰富程度的一项重要指标，也是表明词表对同义词、近义词等有效控制的程度。

等同率高，检索入口就多，可以提高检索效率。

国外叙词表的等同率一般在 40% 到 50% 以上，甚至有些词表的等同率大于 1，即入口词比叙词还多。

我国叙词表的词汇等同率在 15% 左右，叙词表的词汇等同率呈增长趋势。

3) 参照度 (Accessibility Measure)

是指叙词表中每个叙词接受参照项(包括分项、属项和参项)的平均数。

$$\text{参照度} = \frac{\text{分项词数} + \text{属项词数} + \text{参项词数}}{\text{词表中叙词总数}}$$

参照度理想值在 2-5 之间

4) 关联比 (Connectedness Ratio)

是指词表中至少设有一个分项、属项或参项参照的叙词与词表中总词数的比率。关联比和参照度可用于测试词表中词汇之间的连接性能。

$$\text{关联比} = \frac{\text{叙词总数} - \text{无关联词总数}}{\text{词表中叙词总数}}$$

词表中无关联词越少，关联比就越高，词表的性能就越好。

5) 先组度

即叙词的先组程度。可以用词组性叙词占叙词总数的比例计算。

$$\text{pred} = \frac{(a \times 1) + (b \times 2) + \dots + (n \times n)}{a + b + \dots + n}$$

式中，a, b, ..., n 分别表示一字词，二字词，……n 字词。

据兰开斯特的资料，英语和法语叙词表先组度为 1.5—2.0，德语为 1.1—1.2 之间。为保证标引的正确性，避免组配误差，叙词表应保持一定比例的先组词。

6) 范畴和词族规模

即叙词表每一个范畴或词族所包含的词量和等级深度。

兰开斯特认为每一个范畴收词 30—40 个较为理想。

范畴和词族的等级深度则取决于系统显示编制的质量。

从理论上说，范畴和词族的等级层次应专深一些。

但要注意类目收词均衡问题。

七、国内外主要主题词表简介

《美国国会图书馆标题表》《医学图书馆标题表》《汉语主题词表》《中国分类主题词表》《中国社科检索词表》

1. 《汉语主题词表》

(1) 编制

我国第一部大型的综合性的叙词表,由中国科技信息研究所和北京图书馆负责主持,1975年开始编制,1980年正式出版。1991年对自科部分进行了修订,1994年,编制《中分表》,相当于对社科部分进行了修订。1996年,补充了自科部分的轮排索引。

★ (2) 结构

全表收录主题词 108568 个。其中正式主题词 91158 个,非正式主题词 17410 个,词族数 3707 个,一级范畴数 58 个,二级 674 个,三级 1080 个。

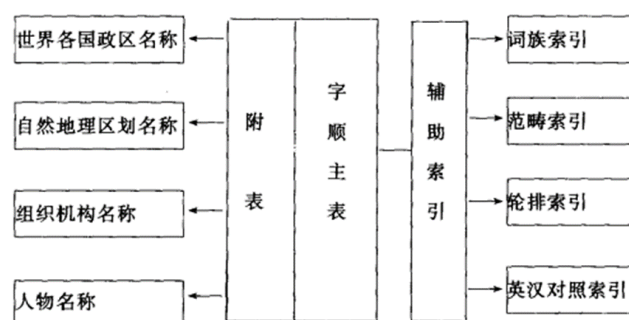


图 4—8 《汉表》组成结构

第 1 卷: 社会科学部分

第一分册是主表(字顺表) A--Z,

第二分册是索引(词族索引、范畴索引、英汉对照索引)。

第 2 卷: 自然科学部分

1-4 分册是主表(字顺表),

第 5 分册是词族索引 A--Z, 第 6 分册是范畴索引, 第 7 分册是英汉对照索引 A--Z。

第 3 卷: 合用的附表

世界各国政区名称表、自然地理区划名称表、组织机构表、人物名称表及英汉对照索引。

★ (3) 《汉语主题词表》的意义

①结构完备。词汇控制严格,整体功能完备,是传统词表编制的一种典型模式。

②词汇丰富。首次为我国文献情报单位提供了一个权威的词汇集合。

③探索了词表编制方法。

④探索了叙词表的实际使用方法。

⑤成为国内文献单位通用的综合性主题标引工具。

★ (4) 《汉语主题词表》的不足

①结构上,篇幅巨大,组成分散,编排不紧凑,整体性较差。

②收词数量不平衡。

③专指度过高。

④词间关系处理上,等同率低于国内外同期专业词表平均值,基本上不设与组代词的等同关系项,无关联词数量大等,是影响词表易用性的比较突出的不足。

⑤《汉表》没有设置统一的管理机构,未确定明确的修订方针,缺乏健全的管理机制:共而影响了词表的发展和使用。

第四章 分类主题一体化语言

一、分类主题一体化词表

1. 分类主题一体化语言的由来及发展

(1) 分类法和主题法在发展中相互渗透和相互融合

- ①分类法借用主题法：编制分类表字顺索引，在分类表中按主题事物集中列类，对类名的词形或词义的控制
- ②主题法借用分类法：在标题表中设置副标题和倒置标题，设置“隐蔽的分类体系”参照系统，在叙词表中引入概念成族和字面成族的方法，编制范畴索引、词族索引和轮排索引

(2) 分类法主题法一体化词表本质上就是一种实现了分类语言和主题语言兼容互换的系统。

	分类法	主题法
术语	类名	叙词
标记	分类号	范畴号
参照	分类参见	参照系统
索引	字顺索引	范畴索引、词族索引等

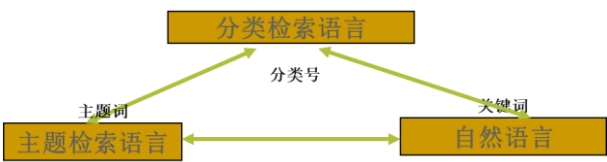
※2. 分类主题一体化语言的原理

(1) 分类主题一体化词表的原理

分类主题一体化语言：在一个检索语言系统中(或由两种原来独立的检索语言合成的系统)，对它们的分类表部分与主题词表部分的术语、参照、标识、索引四部分实施统一的控制，从而能够满足分类标引与主题标引的需要，简称一体化检索语言或一体化词表，国外称为分类法/叙词表系统。

(2) 分类法和主题法在构成原理上是相同的

- ①都使用了分类方法。分类法是显露的学科分类，主题法是隐含的事物分类，聚类的方式也有所不同。
- ②从基本构件来看，实质是相同的。分类法的分类号和主题法的主题词本质上都是主题概念的标识，只不过标识的形式不同，通过一定方式，二者是可以实现转换的。



③类集的内容是相同的。处理的对象都是主题概念，都能够提供概念检索的功能，差异仅在于类集的方式方法有所不同。

※3. 分类主题一体化词表的类型

※	分面叙词表	分类主题词表	集成词表
介绍	通常由一部分面或半分面的分类表和一部	通常由分类号与主题词对应表、主题词与	将某些特定主题领域的若干叙词表和分类

	字顺叙词表组成，有的还附有轮排索引及英汉对照索引。这两部分的类目和主题词一一对应，等值兼容。字顺主题词表起着分类表字顺索引的作用；分类表起着传统主题词表范畴索引和词族索引的作用，可直接用于主题标引与分类标引。	分类号对应表两部分组成。前者为每个类目列出其对应的一个或多个主题词或词串，后者为每个主题词列出其对应的一个或多个分类号。功能上超过分类表+主题词表的总和，成为不同检索语言兼容互换的重要工具。	法汇编而成的一种词表，可以用于联合分类标引和主题标引，用于实现分类法和叙词表之间的兼容及互换。
编制难度	高	低	低
兼容程度	高	低，使用广	低
词表个数	2 个	2 个	3 个及以上
案例	《教育主题词表》、《社会科学检索词表》、《音像资料叙词表》	《中分表》、《中图法教育专业分类法》	《中国图书馆分类法 R 类与医学主题词表 MeSH、中医药学主题词表对照表》、UMLS

4. 分类主题一体化词表的特点和功能

(1) 分类主题一体化词表的特点

- ①在编制方面，采用系统编表的方法，实现词表编制的自动化，可以提高编表的速度和质量；
- ②在管理方面，检索语言的管理工作可由一个机构统一进行，而不再需要由二、三个机构来分别维护；
- ③在标引方面，标引人员可以同时完成文献的分类标引和主题标引，两种标引数据可以互相转换，从而节省人力和物力；成为不同检索语言之间兼容互换的工具
- ④在检索方面，用户可以在一个检索系统中同时进行字顺主题查询和系统分类查询，可以提高检索效率；

(2) 分类主题一体化词表的功能

- ①可以用于分类标引和检索、主题标引和检索以及分类标引数据与主题标引数据之间的转换，用以降低标引难度，提高标引速度，节省标引成本。
- ②可以为进行过分类标引、而未进行主题标引的书目数据库通过主题词与分类号的转换而提供分类检索。同样，可以为上述数据库进行回溯主题标引提供方便。
- ③在这种一体化词表中，加入关键词、自由词等自然语言接口，就可以实现分类号(类名词)、主题词、关键词三者标引和检索的一体化，为它们之间的兼容互换提供工具。
- ④以这种一体化词表为基础，可以建成用于自然语言检索的后控制词表。
- ⑤基于这种一体化词表，利用分类检索、主题检索、关键词检索三者一体化的原理，可以实现自动标引、自动分类等智能信息处理。

二、分面叙词表

1. 分面叙词表的由来及发展

☆2. 分面叙词表宏观结构的改造

- ①详细的分类表代替了以往叙词表的范畴索引和词族索引,用字顺叙词表取代了分类法的字顺索引,简化了词表结构,强化了分类显示功能。
- ②分类表包含了所收全部叙词的全部信息,已经成为一体化词表的主体,字顺表成了分类表的派生物。
- ③分类表中的类目与字顺叙词表的主题词一一对应,用分类号联系,相互辅助。
- ④改变了以往传统叙词表一组词设一范畴号的做法,为每一类目(主题词)均配备一个独立的分类号,使分类标记个别化,便于检索。

表 5—1 《教育主题词表》多重等级关系的显示

《教育主题词表》分面分类表	《教育主题词表》字顺叙词表
F · 各类教育 00554	师范教育 FA 01754
C 教育事业	Teacher education
FA · · 师范教育 01754	S 各类教育 00554
C 教师培训 JQ7 00867	F 高等师范教育 00510
进修课程 NT793 01208
(按师范教育特殊性分)	职业师范教育 02667
FA6 · · · 职业师范教育 02667	C 教师培训 00867
D 职业技术师范教育	进修课程 01208
S 职业教育 FB 02663	职业教育 FB 02663
C 职业学校 GP 02668	Vocation education
	D 技术教育
	职业技术教育
	S 各类教育 00554
	F 职业师范教育 02667
	C
	职业师范教育 FA6 02667
	Vocational teacher education
	D 职业技术师范教育
	S 师范教育 01754
	职业教育 02663
	D 技术教育
	职业技术教育
	S 各类教育 00554
	F 职业师范教育

3. 分面叙词表微观结构的改造

词表的微观结构是指词表的组成单元——款目,即分面叙词表分类表部分的类目和叙词表部分的叙词款目。

(1) 对叙词表构件的改造

- ①增设了词号; ②给每个叙词添加详细的分类号; ③为注释增设专用符号“J”; ④在属项、分项、参项参照之后逐一配备了词号。

(2) 对叙词表构件的改造

- ①新增类级; ②在类名下设置 D 项; ③把类目各种注释纳入新设置的 J 项; ④在类名下设置 S 项、F 项和 C 项参照; ⑤在类名后附加相应的词号,并在 S、F 和 C 项后也附加相应的分类号和词号,以便迅速转查其它类目; ⑥大量设置分面标头(“按……分”),说明类目划分时采用的分类特征。

(3) 类目名称叙词化

- ①分面叙词表通常实行“一个主题一个位置”及“一个主题一个语词”两条原则。

②如果同一个主题在两个学科出现，其含义完全相同，一般分类法对这种具有多重等级关系的主题采用设置交替类目的办法来解决。

4. 分面叙词表的编制

(1) 编制方式的选择

①编制新的分面叙词表。充分利用现有检索语言提供的词汇和词间关系资料；充分考虑与相关检索语言之间的兼容和互换。

②为主题词表引入严格的分类体系。可以把原有的范畴索引和词族索引合并、改造为一个严格的分类体系，也可以直接引入一个分面分类表或半分面分类表。

③把分类表改造成为分面叙词表。这种改造一般使用全分面分类表或半分面分类表才容易实现。把体系分类法改造成一体化词表较为困难，因为它的先组度太高，往往要“动大手术”。

(2) 编制程序和方法

分面叙词表采用了系统方法控制编表过程，编制流程包括：

- ①词表总体设计；
- ②输入数据的搜集和加工；
- ③计算机自动生成分类表、字顺表及轮排索引等；
- ④试标引和修改；
- ⑤编辑定稿、激光照排。

注意：

核心工作：构造分类数据表

基础工作：输入数据，主要包括分面标头、类级、类号、上位词性质、类名及英文译名、类目注释及参照。

编制方法中的优势：全面、系统、事务性工作计算机完成

三、分类主题词表

1. 分类主题词表的由来及发展

2. 分类主题词表的原理

(1) 分类主题词表的原理

基于分类表和主题词表之间的兼容性的转换系统，为分类标引数据与主题标引数据之间的转换提供桥梁。

(2) 分类主题词表与分面叙词表的差别

①分类主题词表一部分等值兼容，一部分在不同程度的兼容，还有一小部分无法兼容；分面叙词表等值兼容。

②分类主题词表是对已经出版的分类表和主题词表寻找兼容关系；分面叙词表是编制一部新表，可以根据需要进行词汇控制。

✖3. 《中国分类主题词表》的结构

(1) 第一卷：分类号—主题词对应表

- ①一部按照《中图法》分类体系组织的体系分类表式的《汉表》范畴索引
- ②一部用《汉表》主题词对类目进行详细注释的新版《中图法》
- ③一部新版的《中图法》。

书：

把《汉表》的全部叙词纳入了《中图法》的分类体系，等于是一部按照《中图法》分类体系组织的体系分类表式的《汉表》范畴索引。它分为左右两栏：左栏的内容与《中图法》第三版的详表完全相同；右栏则是每一类目分别对应的叙词和叙词串。这些对应叙词和叙词串起着类目注释的作用，可以更充分、清晰地揭示《中图法》各级类目的含义，包括类目暗含的各种概念。加上在《中分表》编制过程中对《中图法》第三版 500 多处的增补修订，因此，第一卷又可以看成是一部新版的《中图法》；也可以说是一部用《汉表》叙词对类目进行详细注释的新版《中图法》。

表 3-10 《中分表》第 1 卷分类号—主题词对应表样例

F84 保险	保险
	保险业
F840 保险理论	保险—经济理论
	保险学
F840.3 保险组织和管理	保险业—组织；保险业—经济管理
	保险公司
F840.4 保险业务	保险业务
	保险时；保险费；保险合同；保险金额；理赔
F840.6 各种类型保险	保险—类型
F840.61 劳动保险	劳动保险
	劳动保险费用；[退休金]；失业保险
F840.62 人身保险	人身保险
	人身保险金；旅游保险；人寿保险

(2) 第二卷：主题词—分类号对应表

- ①《汉表》的修订版，增加新词、删除旧词、调整词间关系
- ②分类号代替范畴号，成为一部叙词表式的分类表索引
- ③族首词下的词族全显示，取消了 F 项和 S 项
- ④按照主题词汉语拼音音节排序。

书：

《汉表》自问世以来十多年间未作全面修订。这次在对应表编制过程中，增加了 14000 个反映新学科、新事物的叙词，删除了一批陈旧过时的词，调整了一些叙词的参照关系。因此这一卷可以做《汉表》的修订版使用。

在每个正式叙词后都用对应的《中图法》详细分类号替代了以往的粗略范畴号；另外还编入了大量的与类目对应的叙词串，这一卷相当于一部叙词表式的分类表索引。存在一个主题词和多个分类号对应的情况。

叙词款目的显示内容和方式发生了变化。正式叙词款目保留了 D 项、Z 项、C 项参照，取消了 S 项和 F 项，代之为在族首词款下的词族全显示。叙词词串款目为以往《汉表》所没有的，其中只包括注释项和对应的分类号，不包括任何参照。

改变了《汉表》中的叙词以往按逐字母或逐词排列的做法，一律按照叙词汉语拼音音节排序，并使相同汉字起始的叙词集中在一起，为用户查词提供了方便。

表 3-11 《中分表》主题词—分类号对应表

保鲜剂
TQ459⑦
保险
F84
· 保证保险
· · 保障和赔偿保险
· · 信用保险
· 人身保险
· · 健康保险
· · · 医疗保险
· · · · 养老医疗保险
· · 旅游保险
· · 人寿保险
Z 部门经济
保险—国际组织
F841.3
保险—经济理论
F840; F840.69
保险—类型
F840.6; F840.69

4. 分类主题词表的编制

(1) 手工编制

1) 对应的通用规则

- ①类目对应的主题词必须是词表中的正式主题词
- ②首先选用最专指的主题词对应
- ③若无专指的主题词时，可选用两个或两个以上最邻近的主题词组配对应，优先选用概念交叉组配。
- ④首先列出与类名概念对应的主题词和主题词串
- ⑤然后列出类目注释中的概念对应的主题词和主题词串，以及类目所包含，但未在类名中得到反映的主题词和主题词串。
- ⑥若主题词在分类法中无专指类对应时，一般采用上位类标引或靠类标引的方法，对应于相关类目下。
- ⑦若遇多个主题词与一个类目组配对应时，应依据分面公式：“主体因素—通用因素—位置因素—时间因素—文献类型因素”来确定词序。

2) 一般类目与主题词的对应

根据主题词对应规则，使用主题词和主题词词串加以描述，并将属于该类号范畴内的全部主题词对应在该类号之下。

3) 特殊类目与主题词的对应

- ①类组性类目：对类目中类组类目，各主题概念应分别对应。各主题词之间用分号“;”隔开。
- ②“其他”类目：首先用其上位类目对应的主题词对应；其次是对应其注释中的概念的主题词；
- ③再次是对应其所包含的和无专指类对应的主题词；如果没有“其他”类，则相关的主题词对应入其上位类。
- ④“一般性问题”类目：“一般性问题类目”本身不对应任何主题词。但其下位类对应其相应的主题词。
- ⑤“各种××”“按××分”和带起止符号的类目：此类类目一般以相应主题词加自然语言修饰语方式对应，以三角号“△”为自然语言符号，对自然语言词汇加以标示。

⑥复分表类目：通用复分表、专用复分表的类目均选用相应的主题词对应，其对应形式与一般类目相同。

⑦交替类目：交替类目对应时，一律将所对应的主题词用方括号“[]”括起来，同时用注释指明该主题词对应的使用类目。

⑧组织机构、人物的主题词：有关组织机构、人物的主题词，一般均收入对应表后的组织机构、人物两个附表。

4) 手工编制《中分表》产生的缺点：

①人工对应标引的随意性和不一致性；

②对应标引的主题词有时过于粗略，往往不能标引较深层次的类目含义或隐含主题；

③对应标引的主题词有时过少，有时过多，甚至造成主题词冗余；

④编制费工费时，旷日持久，维护更新周期太长等。

(2) 计算机编制

1) 解决如何剔除标引书目记录中的错误数据

2) 计算机编制方法

①利用现有的书目数据库实体中大量存在的文献分类号和主题词双重标引数据；

②把书目记录中实际存在的分类号与主题词之间的对应关系抽取出来；

③通过浏览、统计、分析和处理大量的机读数据，自动地建立一个具有文献保障的分类法与

④叙词表之间的转换系统；

⑤采用统计学和语料库的方法，来剔除书目记录中的错误数据。

3) 《中分表》计算机编制步骤

4) 计算机编制的意义

①编制的对应款目数量可以超过手工编制几十倍，大大缩短工期，提高对应标引质量；

②直接联通书目数据库进行文献检索，真正实现多种检索语言之间的互换；

③推动自动标引、自动分类以及文献数据库的建设；

④促进我国情报检索系统的发展和网络信息资源的开发和利用。

※5. 《中分表》的功能

①优于国外编制的分类主题词表，功能更强大；

②整体功能强于《中图法》和《汉表》之和；

③降低标引难度，提高标引质量和速度；

④核心的兼容体系；

⑤为《中图法》和《汉表》修订提供依据；

⑥术语学研究和术语库建立提供参考工具。

书：

国外出版的分类法-叙词表对照索引除分类号和叙词外不带其他任何成分，而《中分表》在分类号之下还附带类名、注释，在叙词之下还附带注释和各种参照，成为一个结构完整、可用于分类及主题标引的分类表和叙词表。《中分表》通过对应标引和编辑加工，使两者实现了先组式语言和后组式语言之间的兼容，分类语言与主题语言在结构、语义等方面的兼容，整体功能强于《中图法》和《汉表》之和。《中分表》在《中图法》和《汉表》之间、分类标引数据与主题标引数据之间提供一座桥梁，加快文献标引的速度，提高标引的质量，逐渐形成一个以《中分表》为核心的汉语检索语言的兼容体系。《中分表》对检索语言及其兼容工具的编制产生直接影响，其中包括用这种模式编制专业分类表及叙词表的范畴索引，甚至建立集成词表。为国内自动分类系统和后控制词表系统的研究和开发提供参考工具。

第五章 主题标引

一、主题标引概述

※1. 主题标引的概念

主题标引：用语词（即标题词、叙词、单元词、关键词）来表达文献主题的操作过程，是依据文献内容和价值，赋予文献检索标识的信息表示的过程。按照是否依据词表分为受控标引和自由标引。

2. 分类标引和主题标引

	分类标引	主题标引
标识	分类号（号码）	语词（字、词、句）
标识转换角度	学科	事物（object）
途径	概念—自然语言—分类号 （分类表， classification）	概念—自然语言—主题词 （thesaurus; list of subject headings）

3. 文献信息标引人员的素质

- ①要有较高的文化水平
- ②要具有相关学科的专业知识
- ③要有坚实的文献信息检索理论与实践基础
- ④要充分掌握检索系统的特点
- ⑤标引人员要有认真、踏实的工作作风

※4. 主题标引的方式

- 受控标引：一种依据词表为工具，使用经过控制的语词标识进行标引的标引方法。
- 自由标引：一种不依据词表、由标引人员根据文献主题内容自主拟定标引词的标引方法。

（1）按照标引内容分

- ①整体标引（浅标引）：概括文献基本主题内容的标引。适用于手工检索，一般赋予 2-8 个主题词。
- ②全面标引（深标引）：充分揭示文献论及的所有有检索价值的主题概念的标引。一般使用 5-20 个主题词，组成 3-5 条检索款目。标引深度最大，适用于机检的图书情报部门标引信息价值大的文献，如论文、科技报告等。
- ③重点标引（对口标引）：只揭示文献中适合本专业需要的主题内容的标引。针对性强，不利于文献共享，适用于专业图书情报机构处理相邻专业文献。
- ④补充标引：辅助标引方式，在整体标引的同时，根据特定需要，对文献中部分重要的局部主题进行标引。可以提高整体标引或重点标引的深度。

（2）按照组配方式分：

- ①先组定组式：表达复杂主题的标识是在词表中固定组配好的，标引时只需“对号入座”，

- 直接选取，不必组配。如：《Library of Congress Subject Heading》
- ②先组散组式：表达复杂主题的标识在词表中是单立的，未进行组配，但在标引阶段要根据文献的主题内容予以组配；检索阶段也要通过相同的组配与之匹配。如：《汉语主题词表》《中国分类主题词表》
- ③后组式：指表达复杂主题的标识，在词表中或在标引阶段皆是单立的，不进行组配，检索时则根据检索课题的需要，按一定方式组配查找。如：元词法、叙词法

二、主题标引的流程

查找并利用已有的标引成果
主题分析
主题概念转换
主题标引记录
标引结果审核

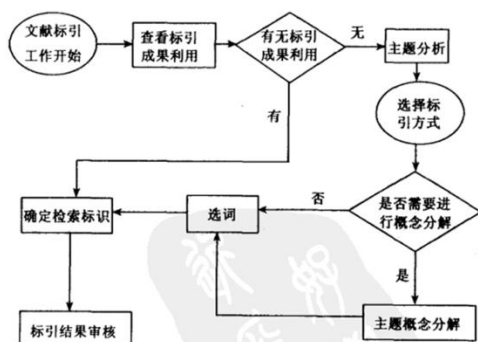


图 5-1 文献主题标引工作流程图

三、主题分析

1. 主题分析的概念

主题分析：根据文献标引和检索的需要，对文献主题内容进行分析，提炼主题概念的过程。要求掌握对文献主题类型、主题结构的基本分析方法。在充分把握文献主题对象及其构成因素的基础上，对主题概念进行提炼和取舍，确定待标概念。

※2. 主题类型

- 单主题：一篇文献中只研究一个中心对象或主题。（只论述一个中心问题或事物）
- 多主题：同时论述两个以上中心问题或事物

（1）按照主题成分

- 单元主题：文献主题只含有一个主题因素
- 复合主题：文献主题含有两个以上主题因素

（2）按照论述的程度

主要主题

次要主题

（3）按照与检索系统的相关程度

专业主题

非专业主题

（4）按照文献论述的直接与否

显性主题

隐性主题

3. 主题结构

主题结构：构成文献的各因素及相互关系。存在多种引用次序（参见第二章）。我国《文献标引规则》规定使用刘湘生的主题分面公式：主体面、通用面、位置面、时间面、文献类型面。

4. 主题概念转换

主题概念转换：在主题分析阶段用自然语言提炼的主题，转换为词表中的语词表达。

（1）主题概念直接转换：词表中有相应的叙词

（2）主题概念分解转换：需要进行概念组配

①交叉关系分解法：外延部分重合，如：企业成本管理＝企业管理＋成本管理

②属种关系分解法：限定成份＋临近的属概念，如：无线电发射接收机电路设计＝无线电发射机＋无线电接收机＋电路设计

四、主题标引的一般规则

1. 总则

①标引的依据：《文献叙词标引规则》 GB/T 3860-1995、《汉语主题词表》、《中国分类主题词表》

②标引的内容：客观地揭示文献所研究或论述的对象的主题概念

③标引的主次：主：文献内容的主要概念和主要对象（主题概念）；次：文献内容的次要概念和次要对象（时间、地点、类型）

④标引一致性：文献主题标目与文献主要分类号的匹配

⑤标引的方式：概括性的整体标引与重点性的分析标引相结合

2. 选词规则

①标引词必须是词表中的正式主题词，不得用非正式主题词标引。

②最切合文献主题的专指词

③可组配表达的词：当词表中不存在专指词时，可以用两个或两个以上叙词（一般是直接上位词）组合形式表达。

④可作上位标引的词或含义相近的相关叙词：当词表中既不存在专指词，也不可能用组配标引时候，可以用上位词标引或靠词标引。

⑤自由词：当专指标引、组配标引、上位标引、靠词标引均不合适时，可以用概念对应的自由词。自由词是词表以外的自然语言。由于词表的更新受限制；词表收入的入口词有限；词表中某类或某些概念收录不全等原因，使用自由词进行补充标引是极为重要的。

3. 组配规则

（1）必须是概念组配

（2）必须概念清楚、确切，只能具有一个含义

组配的次序：主题分面组配公式 主体因素（A）＋通用因素（B）＋位置因素（C）＋

时间因素 (D) + 文献类型因素 (E)

(3) 概念分解——概念组配的准备工作的准备工作

- ①有些概念从字面上分解与概念分解相似
- ②注意有些字面分解和概念分解在字面上相差很远
- ③概念分解不能一步到位

4. 主标题确定的规则

- ①文献内容的主题概念即标引的主要概念，应作为主体因素。属主体因素的主题词均可标引为主标题，也使读者检索的入口词。
- ②当位置因素、文献类型因素成为文献的主要研究对象时，主体因素(主标题)就是这些词。

5. 单主题文献的标引

单主题：一篇文献中只研究一个中心对象或主题。根据文献主题概念的数量和关系，又可分为单元主题和复合主题

(1) 单元主题：不必进行组配。

(2) 复合主题：除从属关系以外，一般应进行组配标引。

- ①交叉关系：应采用交叉组配方式进行标引。
- ②限定关系：应采用限定组配方式进行标引。
- ③应用、影响、比较及相互关系：应进行组配标引。
- ④整体与部分关系：应进行组配标引。
- ⑤事物及其方面关系：应进行组配标引。
- ⑥从属关系：如果同时论述上位主题概念和下位主题概念，应视为多主题进行分组标引。上位主题词和下位主题词之间，不得进行组配标引。如果只是重点论述下位主题概念，则只应标引下位主题概念。

6. 多主题文献的标引

多主题文献：要分解为单主题，再进行分组标引或分组组配标引。并列主题超过四个时，采用上位标引。

7. 各类文献标引规则

- (1) 马列主义、毛泽东思想类
- (2) 历史类文献
- (3) 以人物为研究对象的文献

(4) 涉及地区的文献

- ①以地区为研究对象：用地区名称进行标引；
- ②以地区及其自然和社会现象为研究对象：除从学科主题概念角度标引外，还要以地区名称作为地名主题进行标引。
- ③地区只作为研究对象的区域限定：以其学科内容为主体因素，地区名称作位置因素进行组配标引。
- ④地理类游记：以游记为主体因素，以内容所涉及的地区为位置因素进行组配标引。

(5) 涉及时间的文献

时间因素：事物、对象、问题、现象或状态所处的时间范围，包括年代、时代、时期等时间概念。时间因素通常不作为主标题。朝代、王朝等有明确的时间和地域范围，有独立检索意义，可根据需要将朝代作为主标题。

(6) 涉及文献类型的文献

文献类型因素一般不作为主标目。以某种文献类型作为研究对象时，文献类型作为主标目。文献内容为综合性的，文献类型因素作为主标目。

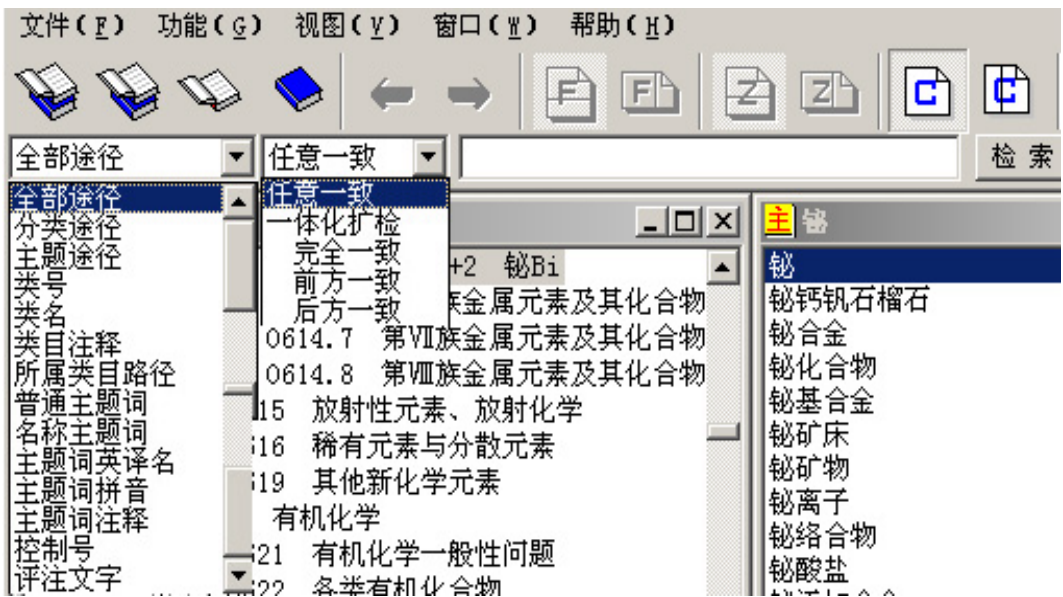
- (7) 文学类文献
- (8) 外语类语言读物的标引
- (9) 丛书、多卷书
- (10) 论文集、选集、全集
- (11) 词典、字典、百科全书、年鉴、手册、目录与索引
- (12) 普及读物的标引

8. 补充：主题标引的思考过程，以《激光在生物学和医学上的应用》为例

- ①查找并利用已有的标引成果
- ②主题分析
判断是单主题/多主题：是多主题，可以分解为“激光在生物学上的应用”和“激光在医学上的应用”两个单主题。
判断是单元主题/复合主题：应用关系是复合主题，采用联结组配标引。
- ③主题概念转换
(入口词的正式主题词 - 最贴合最切合文献主题的主题词 - 可组配表达的词 - 上位词 - 自由词)
“激光”和“应用”拼接成为“激光应用”，提取“生物学”和“医学”无需进行转换。
- ④主题标引记录
可标引为：“激光应用 - 生物学”和“激光应用 - 医学”。
- ⑤标引结果审核

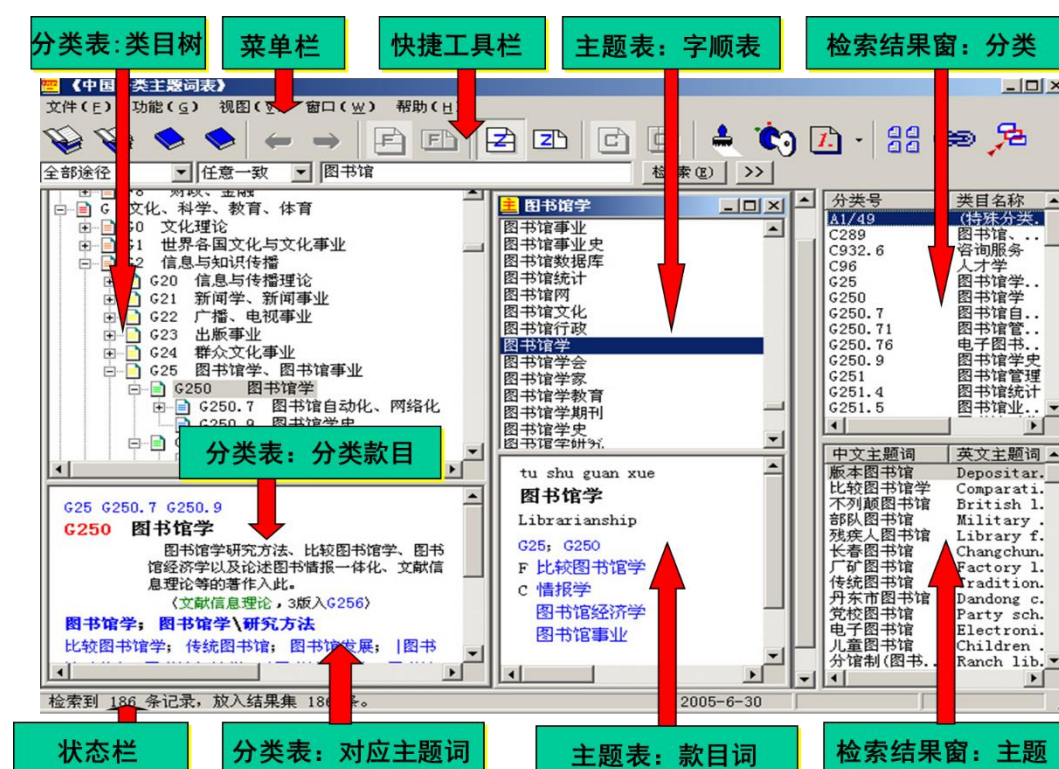
9. 补充：电子版《中分表》

(1) 检索入口



- ①全部一致：自动匹配任意一致，检索范围相当大，适用于宽泛的试探性检索或扩检
- ②分类途径：在分类号、类名、类目注释等子字段检索，适用于排除主题词的检索
- ③类号：一般选择前方一致
- ④类目注释：发现与某主题相关联的类目，适用于分类标引时的比较
- ⑤主题途径：在主题词、主题词注释、主题词译名、主题词拼音等子字段检索，适用于排除分类表内容的检索
- ⑥名称主题词：在普通/学科主题词子字段中检索，排除“名称主题词”
- ⑦普通主题词：在名称主题词子字段中检索，包括人名、地名、机构名、文献名等，排除“普通主题词”
- ⑧主题词拼音：适用于难以输入的汉字的检索
- ⑨主题词注释：一般不用

(2) 宏观结构



17 级题型

一、名词辨析 4*4 个

1. 人工语言/受控语言 & 自然语言/非控语言
2. 分类语言/分类法 & 主题语言/主题法
3. 概念 & 语词
4. 专指度 & 先组度
5. 检全率 & 检准率
6. 标题 & 主题
7. 元词 & 叙词 & 关键词
8. 字面组配 & 概念组配
9. 交叉组配 & 限定组配
10. 范畴索引 & 词族索引
11. 经验法 & 小组汇编法
12. 分面叙词表 & 分类主题词表 & 集成词表
13. 受控标引 & 自由标引
14. 整体标引 & 全面标引
15. 重点标引 & 补充标引
16. 先组式标引 & 后组式标引
17. 单主题 & 多主题
18. 单元主题 & 复合主题

二、简答题 6*9 个

1. 什么是情报检索语言？有什么优势？
2. 提高专指度的措施
3. 降低专指度的方法
4. 提高专指度/先组度对信息检索的影响
5. 引用次序对检索结果的影响
6. 主题法的类型和发展脉络
7. 叙词组配的作用
8. 简述叙词表编制过程中的词汇收集方法
9. 《汉语主题词表》的结构
10. 分类主题一体化语言的原理
11. 《中国分类主题词表》的结构
12. 用户和信息环境如何对情报语言产生影响
13. 情报检索中词汇控制的范围
14. 叙词法为什么比其他主题法语言存在时间长？
15. 评价《汉语主题词表》
16. 如何评价一部叙词表
17. 评价《中国分类主题词表》（为什么 1+1>2；为什么是兼容的核心体系）

三、论述题 10*3 个

17 级考试题目

一、名词辨析 4'

1. 词族索引和范畴索引
2. 关键词和叙词
3. 检全率和检准率
4. 受控标引和自由标引

二、简答题 6'

1. 控制语义关系的方法
2. 专指度对主题标引的影响
3. 相比于其它主题法，叙词法有哪些优点
4. 叙词组配的作用
5. 词汇收集的途径
6. 分面叙词表和分类主题词表的区别和联系
7. 《中国分类主题词表》在信息组织中的作用
8. 单主题和单元主题的区别和联系
9. 写出《习近平治国理政》的主题标引过程

三、论述题 10'

1. 在主题标引中经常会出现哪些错误？如何避免？
2. 设计一个方案实现自然语言和受控语言的转换，并分析这个方案有什么不足。
3. 如何优化《中国分类主题词表》电子版的使用体验。