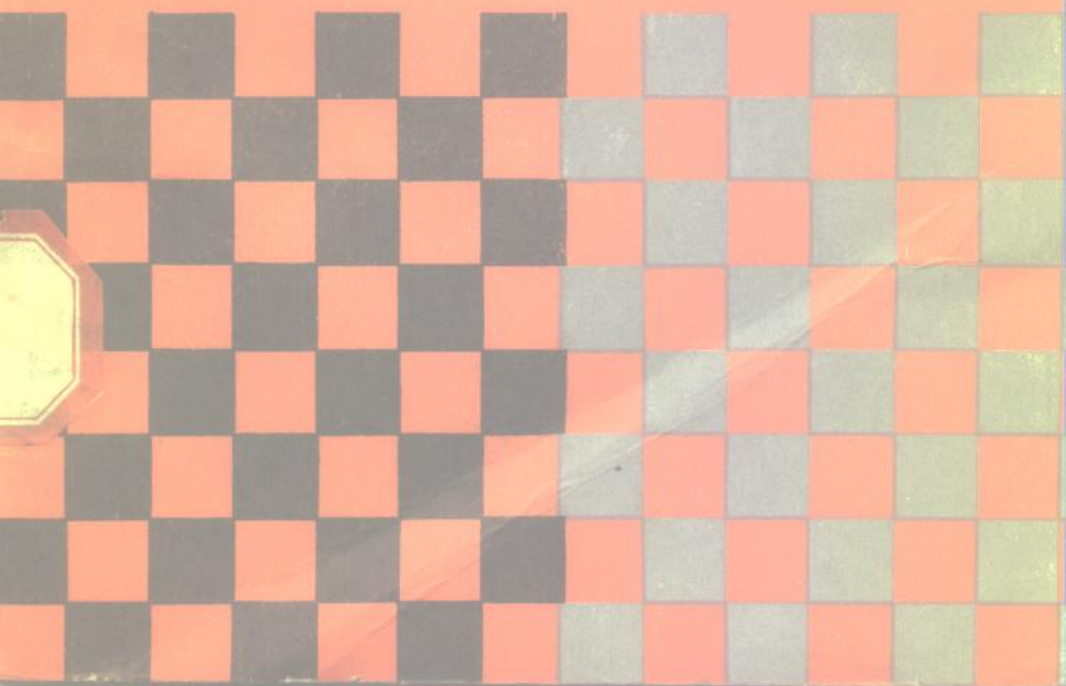


叙词表的 编制和使用 简明教程

〔美〕 F.W. 兰开斯特 著



叙词表的编制和使用

简 明 教 程

〔美〕F.W.兰开斯特 著

赵阳陵 译 钱起霖 校

科学技术文献出版社

内 容 简 介

该书是译自美国情报专家 F.W. 兰开斯特所著的一本讲述如何编制和使用叙词表的压缩课程教材。内容包括：一套共计84张做为直观教具的图解示例和示例解说文字。作者对部分主题做了详细介绍，重点讲述词表的结构和特性。该书内容新颖，图文并茂，简明易懂。

本书适合大专院校图书、情报专业的师生、检索语言研究人员、数据库建设人员和一般情报工作人员学习参考，也可作为培训班的教材。

叙词表的编制和使用

简明教程

[美] F.W. 兰开斯特 著

赵阳 译 钱起霖 校

科学技术文献出版社出版

一二〇二工厂印刷

新华书店北京发行所发行 各地新华书店经售

787×1092毫米 32开本 3.75印张 74千字

1989年8月北京第一版第一次印刷

印数：1—2600册

科技新书目：200—110

ISBN 7-5023-0860-1/Z·122

定价：1.85元

目 录

引言	(5)
单元内容	(5)
第一单元 词表控制的目的	(5)
第二单元 规范化词表的主要组成部分	(10)
第三单元 收集词条	(14)
第四单元 组织词条	(18)
第五单元 族系关系的处理	(21)
第六单元 相关关系的处理	(25)
第七单元 叙词的特性	(26)
第八单元 叙词表的人表款目词	(30)
第九单元 注释和标识词	(32)
第十单元 词表的编排格式和显示	(34)
第十一单元 增补和修订	(40)
第十二单元 计算机的使用	(43)
第十三单元 影响情报系统运行的词表因素	(49)
第十四单元 自然语言系统	(56)
图例	(60)
参考文献	(110)

引 言

本教程源起于阿根廷科技情报中心在联合国教科文组织赞助下，于1978年在布宜诺斯艾利斯举办的“标引语言地区研讨班”。原稿的作者正是该研讨班的主讲人。

研讨班之后，阿根廷科技情报中心根据讲授内容，整理了一份有图例（实际是投影图片）和文字辅助说明部分的教材，并已于1981年用西班牙文出版，题为《标引语言和叙词表的编制》（Curso Sobre Lenguajes de Indización (y) (Construcción de Tesauros, UNESCO出版物 PGI-81/WS/31)。

1984年，作者与联合国教科文组织签定了合同，进行本教程英文版的编写工作。本教材即由此而来。

英文版教程包括两部分主要内容：一套共计84张的图例，这套图例可转制成幻灯片或投影图片；一份对这套图例进行解说和详述的正文。英文版与原西班牙文版差异较大，尤其是英文版本几乎完全着眼于叙词表的编制和特性，而阿根廷研讨班的教材内容则还涉及到其他方面的内容。

正文部分由14个专题单元组成，但是并未试图使所有单元的内容完全匀称起来，因而有些单元相当简要，有些单元的内容则较详尽。

本书在编写时未考虑作为自学教材，尽管它可以为此所用。更确切地讲，选定作为直观教具的图例和指导课程教学的正文说明部分，都是为教员讲课准备的。本书内容并非完

整无缺，如在正文部分各专题中，有许多地方留给讲课教员根据情况自行进行补充(如叙词表和标题表之间的区别)。因此希望教员们能将自己编制的更具体的示例补充到本教程的图例中，特别是对某一专业学生特别感兴趣的具体叙词表的示例。此外，重要的是，应该收集多种类型词表的样例，以供学生们学习时进行处理和实验。同时，较理想的是应该使每个学生都能得到一本联合国教科文组织出版的《单语种叙词表编制与修订规则》(1981年，第二版，以下简称“规则”)作为参考。

最后，还要说明，“词表的编制”课程必须有实习内容部分。要求每一个学生或每一学习小组编制出一个约有200个词条的小型词表。词表在结构方面必须是完整的，即应该包括各种类型的组成部分，如广义词(属)、狭义词(分)，相关词(参)和词条注释等项，这些部分的相互关系应一一对应，而且还应该符合“规则”中各方面的要求。为了帮助学生们进行词表的编制工作，需要给他们提供各种原始资料，如参考工具书、文摘刊物等，以便于学生们从中选取他们所需要的词条。

本书不是课本，为使它的内容相对简明些，对有的主题只做扼要概述。本教程所包括的全部专题及一些相关专题，已在兰开斯特所著《情报检索词汇规范化》(Vocabulary Control for Information Retrieval, Arlington, 情报源出版社，1985年，第二版)一书中作了更详尽的说明。该书采用的叙述顺序与本教程类似。

本教程中提出的要点和列举的示例，大多与联合国教科文组织的“规则”要求一致。但本教程并非旨在用图例仅说

明如何运用这些“规则”，而是包括了词表编制和使用的更为广泛的主题。因而，在文字说明和图例中，有时候也示范了一些与“规则”并不完全一致的名词术语、处理程序和实习内容。这些不同作法目的在于引起对各种编制方法与处理方式的利弊权衡展开课堂讨论。

作者对英国标准协会允许复制《根词表》一书中的有关书页表示感谢。

单元内容

第一单元 词表控制的目的

目的

1. 通过图例表明叙词表在情报检索系统中的作用。
2. 示范词汇不受控制时，会发生的一些问题。
3. 用图例表示出一部规范化词表的主要作用。

图例 1

情报检索 一词一般用来指依照专题需要查出相关文献（书、期刊、文章、报告和其他形式资料）集合的活动。一个**情报检索系统**是由有助于取得有关专题文献的一组活动和实施此种活动的功能设施部分所构成。

图例 1 表明主要的构成部份。从文献普遍获取的意义上讲，由某一情报中心（也可以是一个传统的图书馆）选择和收集那些与一定读者群的兴趣有关的文献。

对选定的资料，要进行著录编目和**主题编目**或作**主题标引**。从概念上讲，进行**主题编目**与进行**主题标引**是相同的。前者通常指对整体出版物（如书、期刊）主题项进行描述，而后者更多用来指处理出版物中的有关部分（如期刊中的文章）。本教程中，**标引**一词指**主题标引**，而所讨论的原则同样

适用于**主题编目**。

标引的第一步是进行**概念分析**，包括确定文献讲的主题。第二步是**翻译**，选用反映文献主题的词。这些词可以称为**标引词**，或可因其描述文献主题，而称为**叙词**^{*}。

一个检索系统使用的一整套标引词可以说是该系统的**词表或标引语言**。

检索系统的核心有两个数据库，一个由文献本身组成，另一个由反映这些文献主题内容的标引词以提供获取文献的线索项所组成。第二个数据库可以看作是第一个数据库的**目录索引**。

对大多数图书馆来说，文献数据库主要由书架上的书组成，至于索引库，则类似卡片目录所包括的著录项。其他单位，索引可能是机读形式（存贮磁带或磁盘）、缩微胶卷或印刷书本式。此外还会有别的形式。

在本图例中，所示从系统用户总体中产生出来的**情报提问**，称作是对该系统的需求。提问者通常是寻找论述一个专门主题的文献或文献的有关部分，而不一定是寻找一个事实问题的答案，当然他们也可以这样做。

在情报中心的工作环境里，接纳情报检索提问与收藏文献处理所遇到的情况相似。其中，**概念分析**是对一个需求者所查寻的目标进行阐明，**翻译**则是通过选用系统叙词表中的词，来描述需求者的需求。对情报需求的描述结果，能够决定一项**查找策略**。

* 关于主题标引准则可以参看国际标准ISO/DIS 5963，文献审读、确定主题和选择标引词的方法（Methods for examining documents, determining their subjects and selecting indexing terms.）

当查找策略用于对索引实施时，那些符合策略要求的索引条目就被检索出来。这种方法可以指引以后从文献库中检索出有关的文献。

由图例 1 中可以明显看出，检索系统使用的词表是它至关重要的一个组成部分，情报需求和文献的主题，都必须用此种语言中的词来描述。

大多数检索系统使用的词表，是一种**规范化词表**。这种词表不过是有限的一群必须由标引员和检索员使用的词。但区别在于，词表规定哪些词可用，哪些词不可用。此外，规范词表通常会有某种“组织结构”（比如，对一些词进行组织来表示重要的相互关系），但这并不是对词所下的一条重要的定义。

在上下文关键词情报检索系统中，与**规范化词表**相反，使用的是**自然语言**。就是说对主题的描述不受限制，可以用文献本身中出现的任意词或词组。关于能否实施一个不用任何词表进行控制的检索系统的问题，将在后面的章节中讨论。但是，不用词表进行控制，会产生图例 2 所示的一些问题。

图 例 2

这里列举的词在某一方面都能与金属的断裂有关。有些是指断裂的形式（破裂、裂开），有些是指能够引起断裂的原因（腐蚀、应力），还有些引起断裂的原因是与金属的特性有关（韧性、强度）等等。

这些词在主题目录或主题索引中，通常是按字母顺序进行排列的，而且它们还要与完全不同课题中的许多其他词散

布在一起。

一个查找有关金属断裂情报的人，会遇到下例中提到的各类问题：

1. 列出的一些词，是同义的或近义的，例如：“裂纹”和“缺陷”。

2. 有些词如脱离上下文，意思就变得模棱两可，如“开裂”(cracking)可认为是断裂的过程或化学工程中使用的的一个步骤（像在石油催化的裂化方面“catalytic cracking”），而象“应力”和“疲劳”这些词能同样用来指人和金属材料。

3. 依照字母顺序，词义相关的词被完全分散排列。而任何人想要进行有关断裂各个方面的查找是不可能的。可能的是某一个人要对断裂的一种形式，如“破裂”方面的全部情报进行检索。遗憾的是，字母顺序排列的结果，把这一课题所涉及的许多词：脆裂、开裂、塑性断裂、韧性、裂缝、断裂、氢脆等等都分开排列了。对一个查找者来说，要考虑到所有这些词是非常困难的，除非用某种方法对这些词进行编组或用参见联系。

图例 3

规范词表试图解决的正是这些类型的问题：

1. 将同义或近义的字或词组进行“合并”，选用一个词，并用某种指引形式将其他词列进词汇表。

2. 多义或在不同上下文中词义有所不同的词，可以多次出现，但使用范围注释或限定修饰词来消除涵义的模棱两可。

3. 用某种方法把涵义相关的词联系起来。图例2中示出一些词的联系。在此介绍两种类型的相互关系。第一种是某物与其所属类（如属和种）之间的永久型关系，破裂和开裂总是属于断裂类的。第二种是瞬变型关系，例如，断裂的过程可以由其他变化过程象生锈或疲劳而引起，也可能是由于受了这种材料本身的特性象强度和脆性的影响。永久型关系有时被叫做是一种**范例或先显**的关系，而较多的瞬变型关系（如断裂可能是由于生锈，但不全是由于生锈，并且生锈不一定总能导致断裂），有时候被视为一种句法结合或**后显**关系。

图例 4

一个情报检索系统中用词表进行控制的目的归结如下：

1. 通过控制（合并）同义词和近义词，区别同形异义词等，促使标引人员和查找者对主题项的描述能达到一致，从而避免有关文献漏检。

2. 用范例或句法的方法把涵义相关的词联系起来，以便进行某一课题的综合性查找。

当使用的词表是规范词表时，标引工作倾向于更加统一。如果用来描述一个课题的词是从事先编好的词表中挑选出来的，标引员会更可能同意选用这些词。查找的准确性也是如此，假若与某一情报需求有关的术语是从给定的词表中选出来的，则这些术语用来针对相应的情报需求就比较容易。据此，规范词表的目的是要把标引员和查找者的语言统一起来。

一部规范词汇表，不过是数量有限的一套术语，它们须

由标引员和查找者使用。然而，更通俗地讲，要确定某种结构，通过某种方法把涵义相关的词组织或连接起来。这是为了帮助标引员和查找者选用最能代表一个专门主题的词和确定进行该主题综合查找而需要的全部术语。

叙词表是一种规范词表，它最通常是用来对反映出出版物内容或个人感兴趣的情报的叙词或标目进行规范化（如进行定题服务SDI的查找策略和兴趣文档）。原则上讲，这样一个工具能够用在有利于术语规范化的任何场合。

第二单元 规范化词表的主要组成部分

目 的

1. 说明一部规范化词表必须有等级部分和字顺部分。
2. 例举几种能够说明这些组成部分的表现形式。
3. 用图例展示一个传统词表的结构形式。
4. 用实例说明词表包含了一个完整的等级部分。

图 例 5

一部规范化词表应该有两个互为补充部分：词的系统（等级）表现形式和字顺表现形式。

图 例 6

这里介绍一个由描述图书馆类型的词群构成的等级部分。它用族系谱学中使用的族系树结构来表示，以便容易看出初级大学图书馆是学院图书馆的下位词，而学院图书馆又

是**高等院校图书馆**的下位词。细分是严格按属种进行的。初级大学图书馆归属学院图书馆类，学院图书馆归属高等院校图书馆类。

图 例 7

族系树能很好地展示相当小的词群之内的等级形式，而非常大的词群之内的等级形式，很难用这种方式表示。这是因为要在一张单页纸上示出全部词的等级是不可能的。而且，族系树的展示法将会浪费版面。

本图例介绍了一个与上例完全相同的词等级关系。但是，词的排列形式与图书馆使用的传统分类法的排列结构相似（如国际十进分类法或美国国会图书馆分类法）。下分各级以行首缩格的方式，表明其等级级别。

图 例 8

本图例看上去甚至更象分类法中类目的结构，因为每个词现在已经给定了**分类号**。这些分类号好象可看作是分类法的一种**标识**。图书馆专业的一些学生对此不了解。在他们看来，分类号标识是属于分类法的基本构成部分。情况不是这样。图例 8 不比图例 7，或更确切说不比图例 6 更表明是分类结构。从图例 6 和图例 7 的表现形式来看，它们是真正的包含“图书馆”这个专业术语的分类法结构。标识仅用于保持等级结构的顺序和为每个术语提供一个方便、简单的识别符号。比如，使用这些符号，就可以组织图书馆藏书的排架。

图例 9

图例 6—8 中词的排列是以体系（等级）方法，而不是字顺方法表现的。这些表现形式必须以某种类型的字顺索引作为辅助，以便向用户说明每一个词在等级体系表现形式中所处的位置。本图例是一个在图例 8 等级排列基础上，编制成简单字顺索引的示例。当然，如果等级表现部分不用标记符号（如图例 6 和图例 7），则字顺索引内可指明每一个词所属族系部分所在的页码即可。这个示例说明了等级结构中不设标识的一个不利因素。

图例 10

图例 6—8 中的词群，主要按等级体系进行编排，而字顺索引实际上是一种辅助部分，即使它作为辅助部分，对各种规模等级体系的有效使用都很关键。一个等级结构有两个独立而又相互有关的组成部分，就是等级结构显示部分和字顺显示部分。

图例 10 选择了同一群有关“图书馆”的词进行展示。显然看出，本图例是按字母顺序排列的。在图例 6—8 中，因等级关系已经明显地通过参见方法编进字顺组织结构中。本例的“属”作为**广义词**(BT)参见，种为**狭义词**(NT)参见。这样，我们能够看到高等院校图书馆一词有两个种和一个属，而图书馆一词有 4 个种，没有属（即，图书馆一词排在树形等级的顶端）。

图例 10 是一个相当简化的图例，说明在一个情报检索词表中词的显示形式。词表与前面描述过的等级结构组合的区

别有以下几点:

(1) 词按字顺排, 实际上包含有等级关系。

(2) 用一种单一序列, 即表现字母顺序又体现了等级结构组成部分。

(3) 在每一个款目上, 只允许反映上下一级的等级形式。这样能够由“高等院校图书馆”一词看起, 看出学院图书馆排列在等级的下一级, 而且再通过学院图书馆的款目才能看见一个等级体系中较低一级的初级大学图书馆。这种“上下一级”的显示特点在词表中并不很重要。但当检索工具以印刷本形式体现时, 为要节省版面, 这个特点就显得重要了。

(4) 广义词(BT)和狭义词(NT)之间的关系应当是对应的。如果词B有NT词Y, 那么词Y就要表示其BT词为B。

尽管用的是掩饰的方法, 图例10中的等级结构显示正象图例6—8中公开展示的等级那样, 同样完整和正确。更确切地讲, 图例10是直接由图例7演变而来的。

另外, 由于图例6、图例7、图例8和图例10都使用的是相同的等级结构, 所以它们的展示便能够自动地从一种转化为另一种。这就是说, 一个计算机程序可以写成将图例10从图例6、图例7或图例8(由于有标识符号做起来更容易)中产生, 也可以写成图例6和图例7是从图例10引伸出来。图例8也能够自动地从图例10中派生出来。然而, 这样做时, 程序需要按规则算法扩展到赋予标识符号。

要注意到, 情报检索中使用的词表完全不同于以罗瑞词表(Roget)作范例的文学术语词表。而且它与图书馆传统使用的标题表也不同。标题表中不具有一套完整的等级结构体