

计算机 辅助术语工作译文集

全国术语标准化技术委员会
计算机辅助术语工作分委员会

编译

语 文 出 版 社

序

计算机辅助术语工作 (computer-aided terminology work), 尤其是术语数据库 (Terminology data bank) 的开发, 自 60 年代以来得到了国际上许多机构和组织的广泛重视, 国际标准化组织术语学 (原则与协调) 技术委员会 (ISO / TC37) 专门下设了一个分技术委员会——计算机辅助术语工作分委员会 (ISO / TC37 / SC3), 负责该领域内标准化理论、原则和方法的研究, 有关国际标准的制订和各项协调工作。

为了适应科学技术飞速发展和术语信息处理自动化的需要, 经国家技术监督局 (ISO 国内总归口单位) 批准, 全国术语标准化技术委员会 (对口 ISO / TC37 相应的国内机构) 计算机辅助术语工作分技术委员会于 1990 年 3 月正式成立, 该分委员会挂靠在国家语言文字工作委员会, 在国际上与 ISO / TC37 / SC3 对口, 国内受全国术语标准化技术委员会领导。委员会由术语学、辞书编纂学、计算机、信息处理技术、情报文献、机器翻译、电子辞典、计算语言学 and 标准化等方面的 20 余位专家组成。

本委员会成立以来, 在各位委员及有关单位的大力支持与配合下, 开展了大量的工作:

——制订完成了两项国家标准: GB13725《建立术语数据库的一般原则与方法》; GB13726《术语与辞书条目的记录交换用磁带格式》。

——正在制订三项国家标准:《术语数据库开发规范》;《计算机辅助术语工作技术要求与评价》;《术语数据库开发用文件编制指南》。

——协助全国术语标准化技术委员会于 1991 年在京召开了“术语学, 标准化与技术传播 (TSTT'91)”国际学术会议, 为会议组织和提供了大量有关计算机辅助术语工作的论文。

——组织国内专家积极参与 ISO / TC37 / SC3 活动, 包括出席有关会议, 编译有关文献资料, 并向其提供我国的有关情况、信息等。

本书的出版于 1990 年下半年列入了委员会工作计划。本书能得以出版, 应该感谢国家科委基础研究高技术司副司长冯思健, 全国术语标准化技术委员会主任易昌惠, 语文出版社社长兼总编辑李行健等的大力支持与协助, 还应感谢各位参与翻译、审校和其他工作的委员和专家们。

我希望能够通过本书的出版, 吸取国外的经验, 在国内同行们和有关部门的配合和努力下, 在该领域内作出更大的成绩, 并为建立具有中国特色的术语学派作出贡献。

全国术语标准化技术委员会
计算机辅助术语工作分委员会
主任 曹先擢
1992 年 8 月

前 言

计算机辅助术语工作是科技术语标准化的一个重要方面，是信息处理、信息管理工作的组成部分。它包括术语数据处理的基础理论、标准化原则与方法的研究，有关标准的制订；术语数据库的研究与开发等。还涉及数据库技术、电子辞典、机器翻译与机助翻译、专家系统与知识库系统和其他信息处理技术。

计算机辅助术语工作在我国是一项开创性工作。为全面了解和掌握国际上该领域的发展状况，在语文出版社的大力支持下，本书得以正式出版，这里谨向他们表示谢意。

本书的编译出版工作由本分委员会主任、国家语委副主任曹先擢主持，原文稿件由全国术语标准化技术委员会秘书处金万平选择和提供，全国术语标准化技术委员会秘书处和本分委员会秘书处联合负责组织翻译审校等工作。参加本书编译和审校工作的有（按音排序）：陈道泉、陈敏、邓洁、丁东江、董柏玉、高青、龚彦如、顾跃挺、郭龙生、郝鸣藏、胡涵景、季光、居绍一、冀立新、金万平、柯志勇、李竹、刘春科、刘庚余、刘宏升、全如咸、荣毓敏、粟武宾、隋京红、田野、万长珍、王惠临、王人龙、王渝丽、王玉梅、吴俭、奚大华、严卫东、詹焱、张浩、郑振声、周剑波、周洁、周智佑等，在此向他们表示感谢。

我们希望本书的出版和发行能推动和促进我国的计算机辅助术语工作及相关领域的各项研究和发展。

全国术语标准化技术委员会
计算机辅助术语工作分委员会
1992年8月

团结协作，大力做好
术语标准化工作。

曹光耀

目 录

序

前言

术语工作与信息管理

术语——信息管理不可缺少的工具	(1)
术语文献工作——公共和私有机构信息管理的工具	(10)
知识管理的关键因素——知识的组织和知识的检索	(13)
信息系统中的认知结构	(16)

术语数据库的理论与实践

语言自动机及其信息库设计的原则	(23)
术语数据库的分类体系	(28)
知识库中的分类体系及自动更新问题	(34)
术语数据库建设和运行费用的若干见解	(37)
奥地利标准的术语数据库	(44)
DANTERM / ORACLE 应用程序——丹麦术语库(DANTERM)概况	(51)
TERMIUM 中语言数据的结构	(67)
运用微机进行术语工作	(74)
TERM-PC——一个运行在微机上的术语库软件包	(85)

文本理解与超文本方法

术语信息处理的超文本方法	(90)
如何把超文本纳入术语库	(95)
文本理解——从书面文本中获取知识	(101)
快速了解不熟悉的领域	(106)

叙词表与叙词库

开发知识库的源泉——情报检索语言	(111)
术语和叙词表——数据库与文本中的知识	(115)
微机叙词表软件的比较研究	(119)
关系数据库环境下事实检索用的联机叙词表	(131)
一种灵活的多语种叙词表	(138)
用 CDS / ISIS 建立国际十进位分类(UDC)数据库	(141)

机器翻译、电子词典、知识库与术语支持

机助翻译与机器翻译中的自动术语支持	(148)
机助翻译工作站软件工程的需求分析、经验研究和建模	(156)
用于机助翻译和科技写作的术语学与词汇学知识	(162)
作为知识获取工具的术语电子词典	(169)

自然语言咨询系统 WISBER 中的术语推理方法	(173)
基于知识的术语框架在 EUOTRA 系统中的应用	(178)
科技文献中的非中性语辞单位及其人工翻译和机器翻译	(185)

数据库技术

数据库硬件和软件评述	(189)
选择数据库管理系统 (DBMS) 的 13 条准则	(195)
数据库工作中的数据重组	(206)
数据库软件目前的发展趋势	(214)

术语——信息管理不可缺少的工具

Alan Melby

术语是信息管理中不可缺少的工具，但是，和所有工具一样，必须正确地使用它。这包括从普通词汇中辨识出术语及使用计算机系统正确地管理术语。专业术语和普通词汇具有不同的性质。一个公司或一个专业领域不能控制和管理普通词汇，但是每一个机构必须控制和管理属于它本身的专业术语，以便产生高质量的专业文献。微机是管理术语、补充主机术语库极好的工作台，并且比起卡片文件来有三个优点：增加了检索功能、自动化管理和更容易共享。但是，建在微机上的术语系统要实现它们的潜力还必须克服一些困难：固定的打印格式、相同文件拷贝间的不一致性、备份数据毁坏、缺少培训和软件硬件系统之间的不兼容。

术语是信息管理不可缺少的工具。这在试图不使用任何专业术语而描述与现代技术的任何方面有关的详细信息之后，立即就变得很明显了。例如，想要精确说明一个词处理文献如何能够通过普通的电话线从一台计算机到另一台计算机，并且达到不用重排就可进一步编辑的形式。这样的说明可能将包括像调制解调器、波特速率、ASCII 码、载波声音和误差检测议定书这样的电信术语。然而这样的说明比起需要提供给设计与制造调制解调器的人全部必要信息的说明，还是相对简单和非专业性的，这要求更多的专业术语。

像所有工具一样，应当恰当地使用术语。本文将特别提出，不能把术语与普通词汇混淆，同时应当使用最有效的技术——主机与微机相结合来管理术语。

1. 专业术语与普通词汇

专业术语中的词和普通词汇中的词都是词汇单位的例子，但是它们具有不同的性质。在严格限定的专业领域，如电学里，从实际的电学家的观点看，我们遇到的是这样的术语：“伏特”、“调压变压器”、“不间断电源”和“欧姆”。这些专业术语具有标准的国际定义，他们都与概念相联系。在理论上，一个概念应赋予一个术语，由此严格地导致在每一种语言中的一个等效的翻译，虽然这种理想的程度因为在给定学科领域内跨子域同义词的存在而常常是达不到的。然而一致的专业术语在给定领域的一种语言内或跨语言使用还是非常需要的，因为它促进更好的交流。事实上，一篇文献内术语的一致用法是达到高质量所必不可少的。

另一方面，普通的词汇，比如“跑”、“美”和“食品”等词就没有标准的国际定义。它们是在特定的语言和文化中定义的，而且它们的含义通过类推、隐喻和转喻可以动态扩展。虽然我们把词、它们在普通语言中的定义与概念联系起来，但是在普通词汇中的概念和在专业术语中的概念是根本不同的。普通词汇中的词可以在有关联意义的连续体中通用，这种连续体可以随时扩展。例如，一篇文学作品可能正在描写看起来只比在风中摇曳的树稍高的云并把它称作 spume，尽管 spume 通常是专指液体表面的泡沫的。spume 的这种用法也许不能再次见到，但是词的其他创造性用法却变成标准的了。例如，如果你不考虑地球的自转和公转，那么除了在地震时，股票市场是不做物理运动的，然而把股票平均价值的日常变化描述成上下运动已经变成一种标准说法了。这种“运动”可能只用左和右这样的字眼来描述，正像法国大革命以来对政治派别的描述一样。普通词汇具有了在新的情况

下产生新的意义的能力。词的新的意义，要能持久，必须由通用的原则（类推、隐喻和转喻）和特殊语言模式来激发，但不能提前精确地预测。

普通词汇和专业术语之间的联系是这样的，当一个特定专业领域里的知识扩展和变化时，新的概念就根据在普通词汇中通行的生成原则与术语相联系了。然而，一旦新知识已经稳定和编入典籍，专业术语就和特殊专业概念而不是和普通语言概念相联系了。只要知识没有发生变化，专业概念就比普通语言概念保持更多的精确性和更少受上下文的影响。稳定性和精确性使得术语对那些共同用特殊的方式观察世界的人们间有效地传递信息变得更加重要了。

专业术语和普通词汇之间的区别是重要的。但是这种不同之处并不总是涉及词的固有性质。不如说，它是这样一种不同，即在特定的情况下在特定的文献中一个词汇单位如何起作用。为了说明这一点，可以观察一个给定的并不总是属于某一类也可能属于其他类的表达法。例如“batch”这个词，通常用于普通词汇中（例如“a batch of cookies（一叠饼干）”），但它也作为专业术语用在工业中（例如滚道窑中的“batch processing（批量处理）”）。普通的和专业的这个词都含有同类的分散的大量处理的意思。但是可以进一步类推，使“batch”作为电子数据处理领域中的专业术语使用，表示计算机运行的一种方式（“batch（批式）”对应于“interactive（交互式）”）。另一个例子是词语“boiling point（沸点）”，它可以是一个化学术语，但也可以用比喻的方式作为普通词汇使用，在像“*He reached his boiling point*”这样的句子中表示愤怒。在这种情况下，必须在相同的表达方式里区分出普通词汇的用法和专业术语的用法。

每一种语言中都有一个普通词汇概念的语义网络，同时在每一个相关的领域中也都有一个专业术语概念的语义网络。这些网络共用许多相同的字符串（像字符串“batch”和“batch”），所以它们经常被混淆。但是区分普通的和专业的这两种网络是很重要的。它们在网络里那些节点上是相同的（即普通语言概念和专业概念与其他节点是有关联的）。这些关联包括子集/上位集关系、同位关系和部分—整体关系。

但是这两类概念对世界万物分类的方法是很不相同的。专业的类是精确的并且相当于一个正规集合。另一方面，普通语言的类是以范例（即有代表性的例子）为基础的。结果是，精确的类服从于诸如并和交这样的集合论运算，而范例类更易变动。

普通词汇是所有说某种语言者的共同的知识和社会财产。国家机构和词典可以影响某种语言。但是管理机构不能指望控制某种语言的全部。没有一个特定专业领域里的标准委员会真想试图控制像法语、德语、那伐赫语这样一种语言的所有方面。

可是，专业术语属于一个领域里的所有的人，虽然他们不说同一种语言。在一个界线分明的领域里控制技术术语对标准委员会和商业机构来说是很重要的，因为在专业文献中一致地准确地使用术语对于良好交流是必不可少的。

然而，面向舆论的标准化组织认为，过分标准化会窒息人类的创造性，不仅在术语的使用方面，而且在科学技术革新方面也是这样。这是因为在一个领域内知识的任何实质的变化，都将要求调整那个领域中的术语。否则术语将不再与概念相匹配，在信息传递中的作用就小了。

人们常常说我们生活在信息时代，但是很少认识到专业信息的有效交流依赖于有效的

术语管理。然而这是很容易证明的，只要想像一下一个没有任何文献工作或有文献工作但不一致地使用专业术语的技术人员，试图维护和修理一台复杂的电机设备零件的情形。仅仅使用普通词汇是得不到令人满意的结果的，就像只靠普通词汇写的说明制造一台复杂的装置很困难所证明的那样。

一些与术语中的标准相反的词可能是由于专业术语和普通词汇之间的混淆引起的。如上所述，这两者之间有重要的区别。普通词汇用来产生新的谈话范围，这样就不能把它限制于固定的意义上。另一方面，在稳定领域中的专业术语是那个领域中特定知识的具体表现，因此必须将它限制在固定的意义上。一旦人们懂得标准化了的专业术语不限制其在普通词汇中的创造性，对规范化的术语就应有更多的支持。

2. 专业术语的多层次

在一个活跃的专业领域中，每年必定有新术语出现。在本文中，“领域”用来描述任一类型的专业工作，其中的人们共用一个他们在日常工作中使用的概念集。这样一个领域可以是汽车轮胎工业，这个领域中的人们共用加工橡胶、制作外胎、嵌进钢“带”或其他增强物的某些概念。有时在制造业中，设计零件的工程师、在车间现场制作它们的机械师和销售它们的经销人员对同一对象使用不同的术语。在这种情况下，谈论子领域变得有意义了。新的术语通常是由一个专业领域或子领域内的一个机构中的普通成员而不是由某些中央术语委员会创造的。许多这样的新术语将在那个机构内消亡，但是一些术语将获得更广泛的承认，直到它们最终在该领域的大部分中得到使用。在理论上，每一个领域都应资助一个管理那个领域中已经变得稳定的术语的标准化组织。无论何时增加了新术语，都应核查它对这个领域知识的冲击。因此，一个领域中的标准术语不应仅仅按字母顺序排列，也应将它插入到显示术语间关系的语义网络中。如果这个领域层次的稳定术语的网络是容易理解的，它就能帮助该领域中的每一个人有效地交流。

然而，新的、不太稳定的术语必须由创造它的机构在一个较低层次上管理。要求人们避免使用那些还没有变成标准的新术语是不现实的，因为只有通过使用，新术语才能成为规范的。管理多层次的术语的关键是，只要有可能就用更高层次的术语，比如全领域范围的术语。这样，每一个机构开发一个数据库，它由这个领域里的标准术语组成，并补充对该机构来说是专门的术语。

在术语管理中没有层次集合数目。例如，在整个工程领域中，就有涉及质量、速度、长度和频率计量的共用的术语层次。世界范围内全部物理和工程科学共用科学计量（和工程数学）的概念。每一类工程里，比如机电的、电子的或化学的工程里，都有一个为这个领域中的每一个人共享的概念和术语的稳定集合。在电子工程分支，如超大规模集成电路数字微电子学中，有一个为那个狭小领域中的工程师共用的术语集。在制造超大规模集成电路的某一特定公司，如英特尔公司里，有一个英特尔特殊术语集来补充较高级的术语层次。在英特尔公司内，一个开发新技术的工程组会有它自己的术语，而一个单独的工程师或术语学家最终将给一个新的概念制造一个新的术语。

当然，术语通常要被筛选或消亡。一个“健康的”术语将获得更广泛的承认并向其应用的最高层次逐步提高。一个“病态的”术语，它不遵守一般的规则，“感觉”也不好，最终将被更好的术语取代。显然，基本属于同一概念的两个健康的术语可能分别“诞生”并通过不

同公司的筛选而在一个较高层次上共存，直到其中一个自然获胜或某个标准委员会能够说服每一个人使用这一个或另一个。

如已提到的，最低的层次由新术语组成，这些新术语对一篇文献来说是特殊的而且代表的概念可能是新的。为了避免混乱，作者应该识别这样的新术语，但这要求每一位作者都能很方便地利用全领域范围的文件和该机构的专门术语。

否则，要核实一个术语是新的并且避免有现成的术语可以使用却另造新术语就很困难。很明显，当作者不知道这个术语已经存在于思想中时，该文件必须允许使用者通过它如何与其他术语相配去查找一个术语，这就要对网络中的概念进行浏览。这是人们反对所谓“词汇编辑”的一个理由，在那里，术语仅仅以字母顺序排列，也许在同一行上有一个等效的译名，但是没有这个领域里知识结构的任何说明。

出现在这里的观点的一个含义是，术语，不论在一种或多种语言中，都不属于一个单独的译者或专业作者。相反地它们属于这个领域中的人的“社会”。这个原则是与译者应保持秘密的、个人的术语档案这个过时观念背道而驰的。在全部的时间里，译者和其他作者如果共享术语就将比仅仅保有他们自己的术语有更多的工作保障，因为高质量的文献只能从始终如一地在所有层次上共享术语中产生。可以期望，术语将不仅由相同或不同组织中的作者之间或译者之间共享，而且通过这样的公司达到机构之间的共享，这些公司有眼光认识到只有很少的术语，比如商标和保密的生产过程说明才是真正独有的，而共享将使每一个人从改善信息交流中受益。

3. 术语管理——质量中的一个可控因素

在特定领域里管理信息包括该领域里的术语，因为语言（除图像外）是专业交流的主要工具。正确管理的术语必须界限明确并限制在上下文有关的范围内。在普通词汇中，一个词的意义能由没有数量限制的上下文，包括作者的意图和读者的内心状态决定。在专业术语中，简单地规定领域，常常就将决定一个术语的精确含义。而且没有其他的术语能在那个含义上使用，否则可能产生混乱。这样，一致性就是一个可控的质量因素了。

学好写作包括许多技巧。当然，语法和拼写是基础，同时编排和体裁也很重要。但是如果我们假定一个专业作者被一个机构雇佣时已经具备了这些技巧，那么唯一最重要的控制专业文献质量的因素就是一贯使用准确的术语。这就说明了为什么术语管理对信息管理，不论是单一语言还是多种语言都是这样一种重要的工具。

当然，也许除了分册目录以外，没有完全由专业术语组成的文献。表面上，我们只须控制专业术语，虽然专业文献中不仅需要专业术语也需要普通词汇。然而，重要的是，受控术语与普通词汇的相互作用也起着重要作用。这样，我们也必须涉及专业文献中规范化的用语和搭配。考虑这点的一个途径是把专业术语比做沙石集合体，把普通词汇比做水泥。那么专业文献就像一块由水泥结合的集合体组成的混凝土板。一篇没有普通词汇的专业文献不能结合在一起，就像松散的石头堆成的柱子。好的“水泥”在一篇文献中是很重要的。

在指出了管理专业术语和相关用语的重要性之后，我们转向如何管理术语的问题。

4. 术语管理中的微机

在前几个世纪，术语没有像现在变化得这么快。今天知识的加速扩展要求新的术语管

理方法。在通用数字计算机变得适用之后不久,有识之士就开始用它们管理术语,结果就产生了今天建在主机上的术语库,如EURODICAUTOM库和Termium库。大型中央术语库是用于管理多个领域的标准术语的很好的工具。然而还有在较低层次上管理术语的可供选择的工具,比如在一个领域内由个人或商业机构维护的较小的数据库。

过去十年计算机产业的发展已经开辟了新的可能性,即在术语管理中使用微机。直到最近,微机还没有足够的力量有效地管理大型术语文件。它们还没有威胁到大型术语库,但是现在,它们有能力帮助大型中央术语库中信息的分配。从一个大型术语库中为一个子领域抽取的术语可以被分配到软盘上。整个Termium数据库都已被装入一张CD-ROM只读光盘上。微机为在一个特定领域里管理术语提供了一个很好的工作台。于是一些建在微机上的术语数据库就能“输送”给一个术语库。微机花费不多,而且在过去几年一些建在微机上的术语管理软件包也变得很适用。对任一专门的机构来说,使用一个建在微机上的术语系统也许比只查阅由许多机构共享的大型术语库更有好处。例如在该机构层次上可能有更多的对专门术语的局部控制和对更新数据库的要求有更快的行动。一个微机工作站的响应时间可能比查询一个远程术语库的时间要少。如果局部系统能包括远程术语数据库的相关部分,同时定期更新,那么一个微机系统就明显地优于一个远程系统。

对于确信术语管理重要并且愿意尝试在术语管理中使用微机的任何人来说,产生的问题是:“我如何能选择一个适当的术语管理软件包?”对术语软件包有几项明显的要求。首先,它必须与现用的硬件兼容。第二,它必须支持所需要的语言。第三,它必须有足够的力量处理当前和计划中的文件。

现在我们将考察一些附加要求的特征,这部分可通过两个途径来考察。如果一个机构仍然使用卡片文件(而且这很可能代表大多数术语用户),那么,这部分可被看成是支持术语管理转向微机的理由。如果一个机构已经使用术语管理软件,这部分可被看成主要特征的清单。

5. 计算机文件对于卡片文件的优点

在个人、机构甚至领域层次上,许多术语文件现在仍记录在索引卡片上,存放在盒子或抽屉里。建在计算机上的术语文件必须有优于卡片文件之处,否则就没有理由将卡片文件转变为计算机文件了。本文的余下部分将叙述这样一些优点并且提出需要注意的地方以便在术语管理中促进扩大使用微机。

优点之一: 增强了检索

为讨论起见,让我们假定存在两个并行的系统,一个由大约30,000张卡片组成,另一个由30,000条计算机记录组成。当然,计算机文件的一个主要优势是检索的速度。对于使用同样的卡片文件的几个人来说,每一次存取都要求用户站起来,走到文件处并且查找合适的卡片。而用计算机文件,用户从键盘输入该术语并且在执行查找命令以后一两秒钟内就能得到响应。对正在使用文字处理机的译者或专业作者来说,不用退出和重新启动文字处理机,计算机文件即可存取。

检索的另一方面,主要与译者有关,是将等效译文拷贝到正被书写的文献中。使用卡片文件,用户不仅必须记住卡片上的信息、把卡片带到工作站,随后还要把它放回原位,而且要把它抄在一张纸上。软件包使用户能很容易地从计算机记录中复制信息并把它贴在

正书写的文献上，不必再打印信息。这种检索的一种变型——不仅与译者而且与只懂一种语言的专业作者有关——能迅速检索出几百个通常使用术语的个人缩写的任何一个全称。因此，术语管理软件对任一术语，不论熟悉与否都是有幫助的。

在翻译中，即使一个术语在源语言中是常用的，检索专门的等效译文——即对要求翻译的机构来说是标准的——也是重要的。对许多术语来说，可能有好几个随机构而定的标准等效词。

一张卡片必须实际存在于文件中一个严格地点上，由一个款目词（在词典中叫条头词）决定。款目词通常写在卡片的顶端。在同一个文件里同一张卡片不应有几张副本，因为否则一个更新了而另一个没有更新，造成不一致。除了款目词以外，常常需要使用某些词或词的组成部分去存取卡片。例如，一个双语种卡片将用两种语言的款目词存取。在一个实际卡片文件内，必须把一种语言选作主语言，还必须有按辅助款目词字母顺序排列的相互参照的文件，引导读者参考主卡片。这使得检索要花双倍的时间。一个计算机文件能拥有很多及时将用户引向主记录的相互参照记录。在一个适当编程的计算机系统内，甚至相互参照构成每一个术语的每一个字和通过一个语义或概念网络的快速导向，比如在给定的概念网络内从普通的词进到更专指的词，也变成可能了。

有好的检索软件的计算机文件比卡片文件明显具有很大的检索优点，而这种优点随着记录数的增长变得更加惊人。总之，对一个术语管理软件包的基本检索要求是：

- 快速的响应，即使是在大型文件上；
- 从词处理到术语软件的快速转换；
- 自动地拷贝并贴补词处理文献；
- 便于将个人的缩写扩展为全称；
- 能在记录中选出的信息上建立快速检索辅助关键词。

优点之二：自动处理

由计算机文件提供的一个优点是增加了处理能力。基本的卡片文件处理包括按卡片顶端的主款目词（关键词）排序。在真实卡片文件中，这必须由人工做，而一个人很容易把卡片排错，使后来的检索混乱或者实际上不可能进行。在这方面计算机文件享有极大的优势，因为它们能自动排序，并且没有一个记录必须从文件中临时“抽出”。

另一种常见的文件卡片处理是生成相互参照卡片。这对于在计算机文件中也能实现自动化的卡片文件来说是非常冗长和费时的。在小型计算机文件里，可用关键词顺序查找全部文件而不用准备。在大型文件里，事先生成相互参照文件能极大地加快以后的检索速度。

前两种处理（由辅助关键字对互见条目排序和生成）是已经讨论过的检索功能的基础。文件的联机更新也是一项基本功能，类似于一个人在卡片文件里补充或修改一张卡片。虽然如果多于一个人同时想要更新同一计算机文件比较复杂。一个适用的术语管理软件包也应允许合并两个文件、抽取子集文件和为打印将信息重新格式化。

当从第三方得到专门的术语汇编时，合并两个文件是非常有用的，在经过适当编辑之后，就可以把它补充到文件中。把大量的卡片插入卡片文件是冗长、缓慢和易于出错的。将一个计算机文件合并到另一个文件中可以自动进行，包括一份计算机生成的任何重复术

语记录的报告。

从一个文件中（通过选择记录或选择字段或两者同时选择）抽取一个信息子集是一个重要的功能。子集可能是适用于一个专门领域的那些记录或用于特定打印输出的那些字段。从一个大的卡片文件中抽取子集是如此的慢，以至于常常是不可行的。

一旦选定了一个子集，为打印信息必须予以格式化。在某些情况下，文件能自动格式化，但是，如果对于一项专门的任务，标准重新格式化算法证明是不适合的，可以生成一个文本并使用一个标准文本编辑程序。即使当信息在文本编辑程序中重新格式化，也比从卡片文件里全部打印它要好得多。

比起卡片文件来，更加局部的自动化处理，明显是通过以计算机为基础的术语管理而享有的另一个主要的优点。总之，术语管理软件包的基本处理要求是：

- 支持快速检索（按主关键字和辅助关键字自动排序）；
- 联机更新数据；
- 文件合并功能；
- 自动抽取文件中已经定义好的信息子集；
- 使信息自动格式化以便为打印做好准备。

优点之三：更容易共享

在几个城市中有办公室的庞大机构内使用卡片文件系统几乎是不可想像的。与制作大型卡片文件副本有关的工作量是巨大的。即使能完成，也只有主卡片文件能被更新，即补充或修改。否则将形成几个不同的卡片文件，并且它们几乎不可能一致。最后，中央卡片文件所在地点必须把每一张更新过的卡片的副本发送出去，而这些副本必须在中央卡片文件的每一个远程副本中归档，因为不然的话，在每一次发生变化时都将不得不复制全部卡片文件。

反之，一个计算机文件使得在建筑的几个楼层或者甚至几个城市中分享术语完全可行。避免产生多种不一致的拷贝是必需的，但是计算机文件的多种拷贝的生成非常容易。计算机文件使得多种分享形式成为可能，包括：

- 在局域网络上共享一个只读文件（类似于允许一群人全部查阅同一卡片文件）；
- 在多台独立的计算机上一个文件的多个只读拷贝（类似于在一个文件里复制所有卡片和用塑料覆盖副本，因此只有主文件上的卡片能被书写）；
- 整个文件的打印形式或一个文件的子集（类似于打印出卡片文件的内容）。

现在如果我们假定更新只能由一个人或一个小组在中心地点进行，而在任何其他地方都不能使文件变化，那么计算机比卡片文件有巨大优势，由于它们有允许有效信息共享的能力。任何基本的术语管理软件包应允许主要的术语学家保留一个术语文件的主拷贝并制作这个机构内所有专业作者和译者都可以利用的只读拷贝。

然而不是所有的更新都在中心地点发生，而且通过发出而共享信息的能力不一定包括这样的机制，即支持引入用以更新主文件的建议的过程。按照定义，只读文件不能有变化，所以只读文件用户必须设法独立地记录变化。这样，用基本的软件包，补充、更新和注解可能必须写在纸上或词处理文件里并且提交给主要术语学家审核。这样一种方法不太理想，因为独自写的注解在格式上可能有明显的差异而且有些甚至在它们通向中心局的路

上错位。

总之，在共享中央文件的信息时计算机文件提供了显著的优点，但是有效地管理来自许多来源的术语信息的更新，同时避免不一致性，可能需要编制特殊的程序。

6.要克服的困难

一些术语管理软件包提供大部分——如果不是全部——上面列出的优点，包括相关的实力与弱点。因此，总的说来，计算机文件比起卡片文件已经有了显著的优点，虽然仍然有一些困难妨碍建在微机上的术语系统发挥它们的全部潜能。所有这些困难都能被克服，因此在审核各种系统时就要很好考虑这些困难。

困难之一：缺乏灵活性

卡片文件，除了它们的许多缺点之外，在信息布局方面原本就非常灵活。如果在一个固定字段上的信息比通常的长，那么它可以被延续到下一行。如果一张卡片填满了，可以增加一张后续的卡片。一个附加的注释字段必要时可加到一张卡片上，而无需把这个字段加到每一个卡片上。计算机文件倾向于比卡片文件少一些灵活性，因为它们常常是基于这样一种模式，在其中每一个卡片（或记录）有固定的定长字段数，使计算机文件灵活并非不可能，但这需要由软件卖主在软件系统早期设计阶段进行一些仔细的规划才能使之成为可能。

一个特别困难的问题是，当专业领域发展和网络变化时保持某一个语义网络。这将可能总是需要人类仔细地介入。

困难之二：不一致性

卡片文件很难复制，而且如果一个卡片文件只有一个副本，就不会有它的多种不一致的副本。计算机文件很容易复制，增加了在不同领域里同一主文件有多种不一致的拷贝的机会。这个危险类似于由几个人编辑同一文献而产生的一个词处理文件有多种不一致的副本。术语库只拥有一个可经电信设备进行存取的主拷贝，用来避免这个问题。建在微机上的术语系统可能不能使用联机电信设备。作为代替，人们可能制作主文件的几个只读拷贝并将一个拷贝发送给机构内的每一个局域网，也许使用脱机电信设备、可移动的单芯盒式磁盘或光盘（比如 CD-ROM 或 WORM）。

没有解决的困难是要求进行局部更新而无须把它们写在纸上或不是术语管理系统一部分的词处理文件里。

一个解决办法是特殊的局域网络软件，它使用一个文件的一个拷贝并允许几个人更新它，利用“锁定”软件防止同一卡片在同一时间被几个用户更新并完全防止未经授权的用户更改文件。这个方法可能适用于一起密切合作翻译同一大型文献的高级译者群体。

另一个解决办法——它考虑到局部用户和不在局域网络上的远程用户——包括允许在注解或更新文件内局部更新软件里的某些种类的工作群体特性，该文件与主文件的只读拷贝相连。在这种情况下，实际的主文件不改变，并且文件的更新部分与只读部分结合起来，以至于它们就像一个文件那样起作用。然后把每一个局部更新文件都送给主要术语学家，他编辑建议、更新文件的主（读—写）拷贝并且定期地把更新过的文件的只读拷贝发送出去。这种处理的好处是：

——局部变化看来已经在主文件里发生了，因此在等待从中央术语学家来的更新文

件的时候不用单独写下来和参照这些变化,使更新过程对用户更为方便:

- 变化已经与相应的记录相连,并且处在对主要术语学家更方便的形式中;
- 主要术语学家不会失去对主文件内容的控制;
- 建议更可能到达中心站;
- 决定中心站如何对建议作出反应的过程能部分地自动进行。

困难之三:不能备份数据

众所周知,磁盘上的数据能在几秒钟内因过热、磁场、操作人员的错误、物理损坏而被毁坏。防止这些灾难的唯一途径是不断地做备份并且在任何时候都至少把一个备份与其他备份分开保存。这最少需要三个用于轮换的备份。当然,卡片文件也能烧坏或以其他方式被毁坏。一般说来,如果电子文件不正确地备份,它们比纸卡片更不安全,可是如果它们被正确地备份,它们就会更安全。

一个可能的折衷方案是以电子方式保留主文件,同时不仅在磁盘或磁带上而且在连续形式的卡片上打印出新的或修改过的文件为它做备份。那么,即使主文件与所有备份有毁灭性的损失,仍然可以完好无损地留下一份有价值的术语文件的纸的副本。

困难之四:培训

在纸卡片上书写不需要专门的培训,而使用术语管理软件包却需要培训。无论如何,学习怎样正确地作术语工作和知道在卡片上或计算机键盘上写什么需要实际的培训。

不过,在最初计划期间就应当考虑到培训人员有效地使用一个系统的费用。

困难之五:不兼容性

随着有能力容易地复制计算机文件而带来了硬件和软件系统之间的不兼容的潜在困难。如果一个领域内术语管理的多层次的概念被严格采纳,那么几乎可以肯定将有兼容性的困难,因为不是每一个人都使用同样的硬件和同样的术语管理软件。随着共享的扩大,不兼容性将成为一个更大的问题。一个比较差的解决办法是简单地重新键入从不兼容系统中收到的任何信息。这将肯定成为向卡片文件倒退的一步。

另一个可能的解决办法是设计一个用于共享术语数据文件的中性中间格式和超越其他不兼容系统的词典数据文件。这样的解决办法已经为在磁带上的以主机为基础的术语库之间共享信息而设计出来了。它就是现在叫做 MATER (来自 TC37 的 ISO 6156) 的国际标准化组织 (ISO) 的标准。它的作者的翻译研究小组正在继续进行适用于微机用户的该标准的实施工作。此外还正在考虑的是给予叫做 SGML (来自 TC97 的 ISO 8879) 的另一 ISO 的标准以更大的注意。SGML 是一种通用的标注语言,用来以一种不依赖任何特殊打印机的方式规定结构化文本的格式,而且一个术语文件能被看作一个特殊种类的结构化文本。提出的用于微机系统之间术语文件的交换格式叫做 microMATER。把 SGML 标准的观点加到 MATER 标准的 microMATER 执行中的重要性在于,它将更好地与加上标题的文本处理方向一致。如果每一个术语管理软件包都允许以 microMATER 格式输出和输入,那么术语数据的交换将变得非常方便。

(唐焱译自 *Term Net News* 1989, No.26, p.3-9, 周智佑校)

术语文献工作

——公共和私有机构信息管理的工具

Michael Schaar

0.引言

当今可获得的信息的规模和数量不断增加，这就使私有和公共用户表现出越来越大的兴趣，他们不仅希望能获得更多的信息，也希望能为特定用户按特定标准提供信息。虽然，数据处理技术的进展已使大量数据的存储和“周转”问题成为次要，但为特殊目的选择提供其中部分数据已日益成为一个结构问题，就是说需要一种工具，能区分出个别的但很复杂的数据结构，供有同类需要的其他用户重复使用。

当今“信息管理”常常受到人们极大地赞扬，但这些赞扬的人们既不知道什么是真正的信息，也没意识到信息管理要面向实际，而不仅仅是经理名片上的一个职务头衔。

复杂的问题需要通过复杂的方法来解决，用简单的方法来解决复杂的问题必然要导致不满意的结果。但是，用复杂的方法得出的答案常常却惊人的简单。

信息管理的目的就是在适当的时间、适当的地点提供正确的信息，并以合理的成本完成这一切。

本文旨在说明术语文献工作是所有信息管理工作的基本先决条件。

1.信息是基石

信息管理是连接机构、信息和技术信息处理三者的中心环节。所以，它的出发点总是要提供当前的、可靠的和结构的基本信息。

这种信息用词语、数字、图像或视听数据形式记录下来，当它们以文本、图像、实况或影片形式组成一定的系统结构时，就为人们提供关于某一情境的有意义的信息。

对于每一个机构，信息都是一种高价值、昂贵的原始资料。由于信息量巨大且处理耗费时间，所以取得信息和存储信息是很费钱的。对每位后勤工作人员来说，任何原始资料必须以典型的产品方式来存储和管理，如同从来没人想到用篮子来贮存油，或想浪费这类稀缺的物品。

但是，这种认识常常没有得到搞信息后勤工作人员的普遍接受。他们往往收集了大量非结构化的数据，并且不正确地称之为信息。后来有人需要这种信息时，就会奇怪为什么查找一番却找不到所需求的结果。

在这里，术语文献工作可以为信息这种原始资料提供一种存储结构。系统的文献工作可以创造出所需要的信息库。术语学方法则为信息管理工作提供分类原则。

2.文献工作和术语工作方法

要成功地使用信息管理的每一个系统概念，这要靠把文献工作和术语工作的现代方法结合起来。文献工作方法提供了宏观结构，而术语工作方法则提供微观结构。用这种微观结构可以对存在于环境、事实或个别数据间的相互关系加以分类和排序。

特别当着着眼于复杂信息结构以及在未来的知识系统中管理和检索这些复杂结构等问题时，文献工作和术语工作方法可作为组织信息系统必不可少的方法，并在信息管理的每个

概念意义上都是基础的工具。

3.作为信息系统的机构

每个公司、每个机构都有其结构、环境和目的。例如，某个机构的结构可包括：

- 层级结构；
- 工作流；
- 生产手段；
- 资源。

其环境可包括：

- 市场；
- 机构和关系的法律形式；
- 技术。

其目的可有：

- 生产；
- 服务；
- 管理。

以上仅列举一些最基本因素。每个机构的信息系统应描述该机构在不同组织层次上的一切组成部分。这就是说，每个部门本身就是个信息系统，但却还要把它当作所有其他信息系统的一个组成部分来看待。

比如，邮局包裹部的发货员和同伴一同作弊，从每个包裹里抽出一件物品，留待将来窃为己有。由于这种作弊有一定目的，他的行为影响了运输部门的信息系统。而这种偷窃行为也进而影响了很多信息系统：纸箱进货部、采购部、仓库、运输成本结算部；而如果此行为暴露，人事部也会受到影响。

从这件事的描述可得出明显的结论，即每个信息系统都从属于一个更高层次（或者用更恰当的说法来说）具有不同结构的信息系统。决定当前系统的价值或等级的，只是加工信息存取机制的位置。

4.技术信息处理

所有这些相互作用的信息系统的组成决定了需要有一个高度集中和统一的系统来设计和实现这个概念。信息管理工作正是要在这一点上起到作用：即规定提供信息的标准。

应采用何种方法来提供信息作进一步处理的规定，要考虑以下几个方面：

- 处理的是哪一种信息？
- 信息要如何处理？
- 信息要存储在什么地方？
- 谁可以存取信息？
- 处理信息的目的是什么？

然后就要规定哪些部门应该纳入信息系统，例如：

- 数据处理 / 数据处理机构；
- 文献工作；
- 后勤(采购出售)；