

查找文献的方法与技巧

龚国伟译

〔美〕赖安E·胡佛主编

上

联机检索策略

湖北科学技术出版社

71027
G354.2
88-15
1

联机检索策略

——查找文献的方法与技巧

(上)

龚国伟 译

联机查检一式

湖北科学技术出版社

联机检索策略
——查找文献的方法与技巧
(上)

龚国伟译

湖北科学技术出版社出版 新华书店湖北发行所发行
咸宁市印刷厂印刷

787×1092毫米 16开本 10.75印张 266,000字
1986年1月第1版 1986年1月第1次印刷
印数 1—7,160

统一书号: 15304·92 定价: 2.05元

译者说明

为了普及联机检索策略知识,充分学习和利用国外这项工作的经验,提高我国计算机联机情报检索的技术水平,促进联机检索工作更好地为国民经济建设服务,我们翻译了《联机检索策略(Online Search Strategies)》一书。

虽然在检索策略一词的定义及其研究范围这个问题上,目前国内外都还存在着某些分歧,但大的方向是一致的。笔者比较同意武汉大学陈光祚老师所下的定义,即“所谓检索策略就是在分析情报提问实质的基础上,确定检索途径与检索用词,并明确各词之间的逻辑关系与查找步骤的科学安排。”说得通俗和抽象一点,就是查找情报(文献)的方法与技巧。仅鉴于此,笔者给译本加了一个通俗易懂的副标题——查找文献的方法与技巧。

初学计算机情报检索的同志在阅读本书之前,应先看一些计算机情报检索、特别是有关国际联机检索的一些基本知识介绍的资料,如中国建筑科学研究院建筑情报研究所翻译的《美国DIALOG和ORBIT国际联机检索系统应用手册简编》一书,以及加拿大图书馆学会出版的《图书、情报、信息人员须知联机检索知识》(注:笔者与武大陈光祚老师等合作编译了此书)。对经济、技术情报的联机检索方法与技巧感兴趣的同志,还可以参看笔者编写的《世界经济、技术信息的现代化检索方法》一书。

本书在翻译过程中对极个别的文句和段落(主要是第七章)有所删节。同时,对原书中几处明显的错误和不妥之处进行了修改。

根据实际需要,我们对本书的下册各章采取了编译的形式,补充了近二年出现的新内容,特别是增加了一些便于读者实际开展联机检索服务工作的新参考资料,从而进一步增强了该书的实用性和可读性。

由于译者各方面的水平有限,而且对原书中所介绍的某些内容还缺乏实际经验,加上时间仓促,译文中一定会有许多不足,乃至错误之处,请读者批评指正。

本书在翻译中曾得到国家科委计算机情报检索领导小组技术顾问、武汉大学图书情报学院科技情报系主任陈光祚副教授的指导和帮助。此外,颜岩、王玫、冯桂林、欧新潮、让德云等同志也为本书做了工作,在此一并表示谢意。

龚国伟

一九八五年五月

于武汉水运工程学院

内 容 提 要

本书详尽地叙述了如何有效地利用国际上通行的联机情报检索服务系统去查找文献资料的各种方法与技巧。书中收入并分析了大量的检索实例与图表，是一本实用性和学术性都很强的书。全书分为上、下两册，上册有五章：1. 美国政府文献的检索；2. 化学情报的检索；3. 生物学情报的检索；4. 能源与环境科学情报的检索；5. 社会与行为科学情报的检索。下册有五章：6. 专利情报的检索；7. 法律资料的检索；8. 医药卫生情报的检索；9. 新闻资料的检索；10. 商业与经济情报的检索。书末还附有一些数据库的生产单位和联机情报检索服务系统(公司)的名址目录，并列出了大量的可供读者进一步参考阅读的书目。

本书既可作为图书、情报、信息、档案与计算机专业工作人员的自学资料和实用工作手册，又可作为大专院校图书情报专业和计算机与信息处理专业师生的教学资料，同时还可作为广大科技人员查找文献资料时的参考用书。

D114/37

目 录

译者说明

前 言	1
第一章 美国政府情报的检索	5
§ 1.1 数据库的类型	6
§ 1.2 CIS/INDEX(美国国会情报服务索引)	6
§ 1.3 ASI(美国统计索引)	8
§ 1.4 CRECORD(美国国会议事录)	8
§ 1.5 FEDREG(美国联邦政府纪事文摘)	9
§ 1.6 FEDEX(联邦活动索引)	9
§ 1.7 VOTES(国会议案表决事项)	10
§ 1.8 USGCA(美国政府贸易合同裁决索引)	10
§ 1.9 检索策略	11
§ 1.10 一个有关立法历史的检索实例	13
§ 1.11 政府情报数据库的更新周期	14
第二章 化学情报的检索	32
§ 2.1 化学情报的检索提问	32
§ 2.2 化学术语表(化学命名法)	32
§ 2.3 CAS 登记号及其术语表(命名)系统	35
§ 2.4 CA 化学辞典联机文档	36
§ 2.5 化学文摘社的化学情报系统	40
§ 2.6 DIALOG 系统中的化学情报检索	42
§ 2.7 SDC 系统中的化学情报检索	51
第三章 生物学情报的检索	64
§ 3.1 生物学情报的数据库	65
§ 3.2 标引规则(措施)	69
§ 3.3 选择合适的数据库	76
§ 3.4 跨文档检索策略	77
§ 3.5 制定一般策略应考虑的几个问题	89
第四章 能源与环境科学情报的检索	93
§ 4.1 由能源所产生的环境问题	93
§ 4.2 数据库	95
§ 4.3 美国能源部能源数据库(EDB)	95
§ 4.4 其它主要数据库	99
§ 4.5 专题数据库	102

§ 4.6 专门文献类型的数据库·····	105
§ 4.7 检索策略·····	108
§ 4.8 检索实例·····	114
第五章 社会和行为科学情报的检索·····	131
§ 5.1 社会科学的范围·····	131
§ 5.2 论述社会和行为科学数据库的文献·····	131
§ 5.3 数据库的发展趋势·····	132
§ 5.4 社会科学文献索引(SOCIAL SCISEARCH)——主要数据库·····	132
§ 5.5 社会科学情报检索中难处理的问题·····	137
§ 5.6 难题的指南·····	139
§ 5.7 检索的类型·····	140
§ 5.8 选择数据库进行跨文档检索·····	141
§ 5.9 联机数据库选择辅助文档·····	144
§ 5.10 社会科学检索中用到的规范化词表(受控词表)·····	145
§ 5.11 规范化词表(受控词)与自由文(自由词)在检索上的比较·····	149
§ 5.12 检索实例·····	155

前言

《联机检索策略》是各学科中一些熟练的情报检索人员根据自己的经验编写的一本实用性很强的图书。它将告诉您如何有效地利用书目文献数据库和计算机检索系统。它并不是一本检索基础知识方面的教课书,也不是关于怎样拟定检索策略的普通论文,它主要是检索人员对检索策略所进行的探讨。虽然检索策略包括所有与检索有关的活动,但我们的目的主要是想通过对检索结果的检验来交流检索的经验。本书的每一位作者都至少通晓一个以上的数据库及其检索系统,而且作为本学科的情报专家和联机检索工作者,他们将自己几年来的工作经验进行了总结,从而为我们提供了十分有益的意见并开阔了我们的视野和思路。

本书所论述的绝大部分内容都限定在当前国际上最流行的四个美国主要的书目文献数据库联机检索服务系统:书目检索服务公司的 BRS 系统;洛克希德情报服务公司的 DIALOG 系统;美国国家医学图书馆的 MEDLARS(NLM)系统以及系统发展公司的 SDC (ORBIT)系统的范围内。当然也有少部分内容涉及到其他一些检索系统。

本书假定您是一位熟悉联机检索并至少使用过上述四个检索服务系统中的一个的读者,或者假定您具有图书馆员或情报专家的某些基本技能,而且能够理解书目情报的有关知识。无论您熟悉哪个数据库或检索系统,都可以把它们作为我们研究和评论联机检索策略的基础或出发点。

下面各章作者一般采用在开始论述时,首先简短地介绍一些本学科领域中主要的数据库这样一种基本写作格式,对那些标引规则可能影响到检索策略的专业数据库给予了特别的注意,主要通过实际的检索图例来说明数据库和检索系统的功能,并往往着重于跨文档或多个数据库的检索策略。

联机系统的昨天和今天

书目文献检索系统通过十年多的发展和改进,已经变得十分先进高级。当1972年和1973年,洛克希德情报系统与系统发展公司分别开始他们的 DIALOG 和 ORBIT 服务时,用户操作选择的自由是很有限的,而且可供用户利用的数据库也屈指可数。用户只能简单地输入一串(行)关键词或一组独立的关键词,并且只能用布尔算符来组配它们。在 SDC 的 ORBIT 系统,你甚至不能够确定哪些字段文句关键词能够利用——只能够利用基本索引检索。那时的系统也不具备将在某个数据库中已经执行过的检索策略保留起来,然后转换到其他数据库中重复执行的功能,当需要多个数据库或多个文档检索时,你必须在每一个需要检索的数据库或文档中重复地进行同一个检索步骤的操作。而且那时的数据库也是十分简单的。

今天,由于可供我们利用的数据库已经有相当的数量,而且它们的规模和标引规则也不尽相同,所以对于检索多个数据库或文档(有些课题,为了得到全面的检索结果,必须用同一个检索策略对多个数据库或文档进行检索)就显得更加困难。这些变化会增加检索者操作的错误,并且检索时间和费用也成为值得考虑的问题。为了解决这些问题,检索系统软件的功能发展到能够促进有效的数据库选择和可用最少字符键进行检索。BRS 可提供专门用来帮助读

者选择数据库的文档；DIALOG 和 SDC 则可简单地回答用户如何选择数据库的提问，指出那些最适合用户课题的数据库或文档。联机系统所具有的保留检索功能，则可允许复杂的检索策略在任何时候、任何一个数据库中再次调用，而无需重新键入这些检索语句。定题情报服务(SDI)功能可以自动更新检索，而无需用户经常进行联机。SDC 系统通过它的“SELECT search”功能，实现了跨文档(crossfile)检索的设想。允许你只要建立好一个检索策略，就可以让它在任何待检文档中逐个执行。

用今天的系统可以建立更复杂的策略，保留或存贮的检索式可以包含以前保留过的策略。同样，这些式子还可以作为更进一步扩展检索的一部分而被保留；有时嵌套(括号)多达10层的检索式都可以作为以后执行 SDI 服务的定题需求单而被存贮在主机系统中。带软磁盘存贮器的微型计算机或带磁泡存贮器的智能终端可以将在脱机情况下键入的并修改好的检索式存贮起来，一旦与联机系统主机联上机，只需按一下个别的功能键，就可将检索式按终端线路的最高通讯速率直接发往系统的主机执行。以上这些功能大大地节省了用户检索自己所需情报而必需花费的联机时间。

使用今天的联机系统，你决不可忽略建立检索策略这样一个极其重要的问题。与情报需求者进行讨论，把检索问题概念化以及确定布尔逻辑式，仍然是检索获得成功的关键。

虽然现在联机系统的功能使得检索工作比以前更容易，但是仍然有下面一些你必须遵循的制定检索策略的基本原则。

关于检索策略的一些思考

由于联机用户们的检索要求和目的各不相同，所以要找出一个在每一种场合都能适用的发展检索策略的通用规则是比较困难的。然而，当您实际应用联机检索系统和阅读这本书的时候，将会找到一些编制检索策略的基本技法。我把这些基本技法概括如下：

- (1) 询问并弄清情报用户的检索目的和要求；
- (2) 把检索课题概念化；
- (3) 利用数据库词汇表选词；
- (4) 与系统进行人机对话；
- (5) 充分利用系统所能提供的一切功能。

下面我将详细地讨论这几种方法。

询问并弄清情报用户的检索目的和要求

虽然情报用户只能告诉您他(她)要检索些什么，但从他们对检索课题和打算如何利用检索结果(指非专利文献)的叙述中，您将会得到比查关键词表更大的收获。如果用户提出的检索课题超出了您的专业知识范围，或您不能够完全理解这个课题时，则不要怕用外行话来询问情报用户，以求得一个满意的解释。如果情报用户懂得一些联机系统的工作原理和使用方法，就应该倾听他们关于检索策略表达式的意见。如有可能，最好让情报用户在场与您一同进行检索。如果不行，也应与用户一起仔细检查检索结果，以便总结经验改进检索。

总之，不能够忽略参考询问用户意见的重要性。

将检索课题概念化

分析检索要求，划分出检索课题的组成部分或者说把课题划分成积木块。当然，也有少量的检索课题只涉及到单一的概念。文氏图和布尔逻辑是使检索课题概念化的有用工具。除

非课题要求的是一个综合性的书目,一般,开始时总是先用课题中较专的概念去检索,如果没有命中,再换用上位概念来放宽检索。如果命中文献太多,可采用与其它概念组配来对检索进行限定。采用先检较专概念后检一般概念的目的,主要是为了节省检索时间。因为有时候一开始采用课题中最狭窄的概念组面进行检索所产生的结果,可能恰恰就是用户所需要的。

概念化处理可以应用得很广,一般应将检索课题分解成为许多细小的专门概念,然后将它们进行各种组配来限制检索,以防止命中文献太多或为零。对一些专业数据库的检索尤其如此。例如,您如果直接用“Child abuse(儿童陋习)”这个概念去检索 CHILD ABUSE AND NEGLECT 数据库,就可能没有任何结果(命中为0)。同样,如果您使用“Patent(专利)”这个词去检索 WPI(世界专利索引)数据库,则也是白白地浪费时间。

此外,还应该注意去掉所有与课题要求无关的概念。因为不少参考文献可能同时论述到合乎某个课题所要求的和不需要的东西,频繁地使用 NOT 算符将会把一些同时含有重要概念和无用概念的文献去掉。所以使用 NOT 算符时应特别慎重。

利用数据库词汇表选词

不少数据库都使用规范化词表和(或)分类码表。主要概念一般都规定要使用叙词或代码。因此,在进行联机之前,要利用叙词表和词典的指引来找出恰当的词汇术语及代码。现在不少数据库的词表都已装入主机系统,因此可以利用联机检索辅助功能来进行这一步工作。常用的操作指令有 EXPAND、NEIGHBOR、ROOT、EXPLODE 和 TREE 等。虽然利用联机数据库词表辅助查词会花费几分钟的时间,但却能免遭失败和节省正式检索的时间。联机显示几篇试检出来的文献的篇名和标引(词)字段记录,也可以从中发现一些更恰如其分的检索词条。如果您使用的是发报速率为1200波德(baud)的终端机,则实现这一步工作就特别快和容易。

当数据库有分类码时,可利用它来检索上位概念(主题)。这些代码可以被看作是概念码、主题码、范畴码等等。CA SEARCH(化学文摘数据库)使用的化学化合物登记号可以简化在那些使用该号的数据库中所进行的涉及化学方面的检索。BIOSIS PREVIEWS(美国生物学文摘)则使用了一个分别对应主要概念和有机体的由加权概念代码和生物系统代码组成的完整代码系统。

与系统进行人机对话

联机检索系统一般都被设计成人机对话型。有些检索者主张要获得成功,就要在开机前制定一个完整的检索策略。他们输入一个完整的检索(式)、大量指令,并在按 return(回车与执行命令)键前就输入了脱机打印命令。这不是人机对话式的检索,而是联机批式检索。他们忽略了系统的人机对话功能,其结果可能会出现许多不切合检索课题的东西。当然,如果你特别熟悉所使用的数据库、检索系统和检索课题,也可能会获得成功。但决不能把它作为常用的方法。与系统进行人机对话,可以从了解到不少关于数据库和检索系统方面的各种功能及其优、缺点。

当联机时,如果对检索的结果没有把握,则可从命中的文献记录中抽一、二篇进行联机浏览,以便改进检索式,然后再抽样检查结果,再修改检索式,如此反复地构造和改进检索策略,直到结果令人满意为止。否则,保存您的结果,脱机后仔细检查,然后换用一种不同的方法或途径重新组合和重新开始检索。首先应放宽检索范围并应查询系统。有时候,你可以在不断地与系统进行人机对话的过程中学到一些在联机服务公司出版的用户手册中以及他们所举办的训练班中学不到的检索技巧。

充分利用系统所提供的一切功能

联机服务系统的容量和功率得到不断地增加,其功能也得到不断地改进和完善,检索终端上的打字输入也变得更加简单,系统的响应速度、准确度和效率也不断地得到提高。但一个简单的键入上的错误仍然可能会导致整个检索无效,在一些文档中,一个差错就会付出昂贵的代价。数据库也会呈现印刷上的错误——这比出现在其它地方更坏。这些错误可以通过 EXPANDing 和 NEIGHBORing 指令扩展数据库索引来发现。有时候,某个遗漏的词条可能正是检索所需要引用的。BRS、DIALOG 和 SDC 允许将系统显示的索引中的词条转为检索词或检索语句,而无需用户自己键入它们。

当你检索非规范化的字段(如篇名、文摘等原文字段)时,可利用邻近(相邻度)关系式或“全文”算符以及截断算符,这些功能可以补偿词句拼写的变化和提高检索的精度。为了避免对每一个检索词都要进行限制,一些系统设有限制整个检索语句甚至限制全体检索(在 SDC 系统)或后置限制检索语句(在 BRS 系统)的功能。

在有联机词表的数据库中(DIALOG 系统),可以利用 EXPAND 指令来扩展有关的索引词。使用带 AUDIT ON(SDC 系统)和 SET DETAIL(BRS 系统)的检索语句,可显示词表所列出的所有索引词条。使用 SELECT STEPS(DIALOG 系统)分步骤选词指令可以在组合式中多次重复地利用以前已键入过一次的检索语句(只需简单地键入代表该语句的组号数字即可),而无需重新键入它们。

几乎所有的联机检索系统都具有高效率的保留和存贮检索策略式的功能,可以用于跨文档检索。CROSS(BRS 系统)、DIALINDEX(DIALOG 系统)和 SELECT search(SDC 系统)指令都具有很强的功能。特别是 SDC 的“SELECT search”的跨文档检索功能最强。让计算机系统去为你做那些检索策略逻辑部分所要求的冗长工作吧。

结 束 语

我们虽然在尽可能提早《联机检索策略》一书的出版时间问题上作了多方面的努力,但由于联机情报工业的发展变化是相当迅速的,所以本书中的某些资料可能在出版后已经过时。读者应该经常查阅常用数据库的编制单位和联机检索服务公司所提供的有关检索策略技术的最新资料。

因为作为一本图书,它不可能代替众多的数据库的编制者、联机检索服务部门以及一些经验丰富的检索人员所编写的优秀手册、检索指南、训练教程和论文,而后面这些资料将会使您学到更多的有关数据库、联机系统和检索技术方面的知识。我们希望本书能够引起您的检索兴趣,使您试着去采用以前没有考虑到的新途径和新技术。本书附录列出了一些数据库的生产单位和联机服务系统公司的目录(名址录),并列出了可供您进一步阅读参考的书目。

如果您现在还不是但乐意成为一个联机检索者,或者您打算在你们的图书馆或信息中心增开联机检索服务业务,我推荐您在读本书之前先读我早些时候所编写的《图书情报人员联机检索指南》(知识工业出版公司,1980)。祝您检索顺利愉快!

赖安·E·胡佛

1982年10月

第一章 美国政府情报的检索

海伦·贝尔&谢丽尔·罗森

与美国政府生产和提供的大量情报相比,即使最大和最高级的私营情报收集组织也相形见绌。仅美国国会每年的工作报告就超过了七十万页。它比其他政府机构创造和传播的情报都要多。事实上,美国政府是当今世界上最大的情报生产者和标引者。它的各个部门和机构生产了包括各个学科和主题在内的大量数据库。这些数据库索引和摘录了政府及其承包商的出版物,以及其他一些国营和私营机构的出版物。一些主要的政府数据库,例如 AGRICOLA (农业文献索引)和MEDLINE (美国医学图书馆医学文摘),将在本书的其他章节中加以介绍,本章主要是讨论对美国联邦政府日常工作活动中产生出来的情报所进行的检索。

一些私营公司已专门致力于编制这些由政府部门生产出来的浩如烟海的情报,并将它们进行加工后存贮到计算机的磁盘或磁带存贮器中。下面我们将分别介绍由四个公司编制的七个数据库,如表1.1所示。

表 1.1 美国政府情报数据库

数据库名称	编制机构	内容范围	更新周期	检索服务系统
ASI 美国统计索引	美国国会情报服务公司	美国统计索引;选自联邦政府机构的出版物;1973年~现在。	月刊	ASI, DIALOG, SDC
CIS INDEX 美国国会情报服务索引	美国国会情报服务公司	国会情报服务索引;美国国会的委员会及其下属分会机构;1970年~现在	月刊	ASI, DIALOG, SDC
CRECORD 美国国会议事录文摘	首都服务公司	国会议事录;美国国会的立法活动;1976年~现在。	周刊(在国会会期内)	ASI, DIALOG, SDC
FEDREG 美国联邦政府纪事文摘	首都服务公司	联邦政府纪事文摘;1977~现在	周刊	ASI, DIALOG, SDC
FEDEX 美国联邦活动索引	首都服务公司	联邦索引:国会纪事录、联邦政府纪事录、总统文件每周汇编、贸易商业日报、华盛顿邮报、F&S索引、联邦管理法规、美国法规等;1976~现在	周刊	ASI, DIALOG, SDC
USGCA 美国政府贸易合同裁决索引	华盛顿代办人服务公司	美国合同裁决,资料来自商业日报的A、H、U版;1978年~现在	月刊	SDC
VOTES 国会议案表决事项	政策评论联合公司	美国国会提出的议案案卷;1979~现在	双月刊(在国会会期内)	SDC

§ 1.1 数据库的类型

政府最重要的工作之一就是监督其机构本身的活动,记录国会有关立法和法规的活动及其在各种大小、公私以及国际国内的一些组织机构中的影响作用。几个联机数据库——CIS INDEX (Congressional Information Service Index: 美国国会情报服务索引) CRECORD (Congressional Record Abstracts: 美国国会议事录文摘)、FEDEX (Federal Index: 美国联邦活动索引)、FEDREG (Federal Register Abstracts: 美国联邦政府纪事文摘)以及 VOTES (国会议案表决事项)——都已被编制成为有利于计算机处理的情报形式。

美国联邦政府也生产了不计其数的统计情报。ASI (America Statistics Index: 美国统计索引)索引和摘录了由联邦政府生产的统计文献。

USGCA (United States Government Contract Awards: 美国政府契约合同索引)则包括了所有由联邦政府授权订立的合同通告。

§ 1.2 CIS/INDEX (美国国会情报服务索引)

涉及到政府情报的数据库既有十分复杂的,也有十分简单的。CIS/INDEX (简称为 CIS) 数据库可以说是现有联机系统中比较复杂的文档之一。它索引了大量的国会出版物,并采取了准确严格的标引和摘录措施,确保了整个数据库的一致性和质量。如果要最佳状况地利用该文档,则需经过一段时间的训练。

CIS 数据库提供了国会的 300 多个各种委员会及其下属机构的所有工作报告。负责收集该库文献的工作人员每个星期都要访问一次国会,每个月都要对国会的下属机构进行访问。其文献的类型由一个可检号码来指明。共有如下几种:

0 类——特种文献:包括立法的文件信函、年度的或定期的报告、否决的议案咨文、议会会议休会与延期的公告、听证会的颁布、特别形势报告或特别收集的资料等。

1 类——听证会文献:如立法听证会,研究听证会,对有争议的问题的调查会,失察(政府工作失误)听证会,授权和拨款问题的听证会或批准任命的听证会等。

2 类——委员会的出版物:包括专门的资料文集、特别情况和问题的报告或听证会摘要。

3 类——报告:包括有关立法的报告、特别会议、听证会和委员会报告。

4 类和 5 类——行政报告和公文:包括有关条约和协议的建议、同意批准的报告以及请求批准任命的公文。

CIS 不包括礼仪和国会内务管理方面的文献。

主要记录和分析记录

在以上文献类型中要算听证会资料的记录最长,这些资料被用一种独特的方法加以处理,每一个听证会文献(引文记录)都由主要记录和分析记录二部分组成。主要记录(正文)包括一个完整的听证会引用文献目录再加上听证会的文摘,主记录后面是个别的分析记录,包括与听证会有关的主要证明或旁证材料的文摘及其资料出处等项目。

证明(附录)材料的文摘仅仅只索引了与听证会其他部分材料不同的项目,包括:主要和次要的见证人名字、公司名称、组织和政府机构的代表以及撰写文章的作者及主题等。因为

检索者一般只习惯于看那些作为单独记录出现的题录或引文,所以 CIS 数据库的新用户在看联机打印输出时往往容易被弄糊涂。但只要注意登录号,还是比较容易区分主记录和分析记录的。登录号包括该文献输入 CIS 数据库的年号、发表该文献的委员会、出版的类型再加上一位用于识别每条独立的证明资料的十进制数。例如,有一篇主记录的登录号为:78—S361—24,其中 78 是 1978 年;S 代表参议院(Senate),361 是财政委员会听证会的序号;24 是该委员会听证会的第 24 卷材料。因为分析记录一般被当作是紧跟在主记录后面补充主记录的附录材料,所以该主记录的第一篇证明资料的登录号就被记为 78—S361—24·1。

因 CIS 的方针是力求对记录的描述越准确越好,而不是以记录的简洁为宗旨,其结果使得文档中的一些记录变得十分冗长。针对这种情况,该库设立了表示记录长度的字段,指出作为一个完整的听证会材料的一部分的分析记录的数目、主记录中的字符数目以及主记录加上所有分析记录的总的字符数目等。记住这个关键的特性,对于你使用 PRINT (打印)命令是很重要的。为了避免出错,检索 CIS 文档一般使用 LINK 或(L)算符,而不用布尔算符 AND。LINK 可以确保主记录和分析记录之间的逻辑联系,而不至于使证明材料配错主记录。

索引的词条

CIS 数据库的标引人员从自编的叙词表中选择了最适当的特定款目词或词组,而没有标引太泛或太窄的词条。在印刷本的索引中,CIS 为用户提供了“SEE(见)”或“SEE ALSO(参见)”两种词间的参照形式。联机用户们可以查阅印刷本的《CIS Index》或查阅《CIS Onlin User Guide》与《CIS Thesaurus of Index Terms》。这些印刷本的工具书索引了所有国家级和绝大部分州级组织机构的名字。委员会名字可由数字号码检索;委员会的下属机构名称可以直接用名称检索。如果需要,标引员也可选取文摘中的词来进行标引,如“Bibliographies”、“Glossaries”、“Congressional Committee Activity Reports”以及“Congressional Committee Calendars”等都是很有用的标引词。

CIS 数据库索引了主题词以及议案或法案的名称,普通的名称如同官方的法案名称一样,也被包括在索引中。

凡是主记录中所涉及到的一些有价值的补充资料,都在文摘中给予了特别的说明。如增补的详细图表、特殊格式的数据、非出版类资料、一些不著名的杂志中的文章等。标引员根据个人或团体著者、恰当的主题词以及法案的号码与名称等来标引这些资料。

为了准确和可靠,见证人的名字要仔细核对。尽管力求按证明材料中出现的先后顺序来输入见证人的姓名,但如果见证人超过 15 个,则按其姓名的字母顺序来排列。如果证明材料表示的是同一观点,就将它们集中于一篇分析记录中。

立法报告文摘包括委员会的建议事项、对于法案的修正案、法案的名称与标题、改进法案的做法、将受到法案影响的行政部门或将付诸实施的法案、委员会成员们的观点、立法的前后过程以及主要的措施和决定等。

CIS 数据库提供了大量的内容十分丰富的国会各委员会的情报,这些情报资料通常是由国会研究服务处(CRS)编制的。其中不仅有一些专门的听证会和立法方面的分析与归纳资料,而且还有涉及某个专门课题的研究资料。如,查看最近的一篇约 900 页的委员会出版的资料,其中有关于越南老兵的整顿问题,有银行和金融方面的参考指南(包括一个术语小辞典),以及一篇分析和总结有关 1981 年经济恢复税法案的文章。委员会出版的文摘上报道了出版物内容的提要与目录。

一般, CIS 数据库的标引和摘录工作原则是坚持先查实所有现存的有关数据, 如法案名称与号码、附录材料、图表、著者和引用的资料等。卷帙浩繁的美国国会工作记录文件就是这样被完整地标引和摘录的。

§ 1.3 ASI(美国统计索引)

由美国国会情报服务处(CIS)编制的另一数据库是 ASI(美国统计索引), 与上述文档不同, ASI 并不记录政府的工作程序, 而是索引和摘录联邦政府出版的所有统计资料。

因为该数据库也比较复杂, 在联机检索以前先查阅由 CIS 出版的印刷本的《ASI Index》, 常常是很有好处的。和 CIS 索引数据库一样, ASI 也分主记录和分析记录。主记录是从总体上描述出版物, 包括一般目的、出版物的种类与格式, 重点在于统计数据类型以及主题范围。地方机构的统计情报、部分科技出版物中的数据和图表也包括在主记录中。分析记录收集的常常是一组图表、文献的某些章节、连续出版物的一部分、月度报告或其他定期出版物。

索引的词条选自 CIS 自编的叙词表。单卷本的分类索引对于标引数据库是非常有用的, 而且对于制定检索策略也是必不可少的。这些索引类目可由诸如国家、疾病、产业……这样一些范畴来划分。当文献中含有统计资料时, 可根据并参考该索引中给出的若干例子来选用其上的词条。由于该索引并没有将每一个独立的词条都编入其中, 标引员将用分类代码来表示那些含义比较广的词条。例如, 如果你想检索美国马里兰州麻疹的发病率, 则应该用一个 /CC 加上疾病代码(L)一个州名代码的策略, 而不宜采用 MEASLES LINK MARYLAND。这里再次提醒大家注意 LINK 或(L)算符的运用。

国会文献中的统计情报在 ASI 数据库中也被索引和摘录, 但重点放在统计数据上。

供整个文摘所使用的规范化缩略语词表, 十分清楚地说明了用于标引的缩略语形式。ASI 所标引的主题词一般都是比较重要的数据。如, 专有名词、地名和范畴分类等。而且标引词大都采用专指度高的词, 例如, 某篇记录采用专指的词条“Mine Accidents and Safety(采矿事故与安全)”, 而不用泛指词条“Mines and Mining(矿藏与开采)”来标引。

§ 1.4 CRECORD(美国国会议事录)

经常与 CIS/Index 配合起来检索的数据库是 CRECORD, 它是美国首都服务公司(Capitol Services, Inc. 缩写为 CSI)编制的《Congressional Record(美国国会议事录)》的机读版本。该数据库是国会议程与会议记录的官方报道, 其涉及的内容十分广泛。输入该库的记录是首都服务公司(CSI)的标引员根据日报上的有关报道资料提炼加工而成的题录和文摘。但标引员没有对这些报道加以分析或插入自己的意见。该库主要提供叙述型和报道型的情报, 有适当的作为引证用的图表与实例。虽然简要, 但文摘还是比较详细的。

美国国会议事录包括法案和决议案、议案的修正案、委员会及其分会的报告、政府公法、议员的活动、委员会及其分会听证会的议事日程、演讲、总统或议长发布的行政消息以及议会的议事日程等情报。

用于 CRECORD 的是一个非规范化的词表。而《Congressional Record》中的记录则是根据一个约有 350 个分类码的分类表来组织的。这些分类码都是根据主题来确定的, 该表的类

目分得较粗,足以适用于一个比较宽的主题范围。每一个记录都有一个到二个分类代码与之对应,如果必要,也可用二个以上的代码来标引。标引员通过查阅一本自编的词表来选择分类代码。分类码表比较稳定,很少进行增加、删改与合并。为了完善检索,必要时也可以用新的分类码取代旧的。CSI编写的《CRECORD 用户指南》详细地叙述了分类码及其在改进检索策略中的作用。

CRECORD 文摘包括讨论的法案、投票表决情况的记录、正式通过的修正案、否决以及搁置或撤消的议案、反对意见和支持意见的声明书、图表和其他有关数据。文摘对原文中出现的有关行政部门提出的议案和文件也进行了加工。对于提出的法规、法案的号码、倡议人、共同倡议人(按姓名自由排列)、法案的规定和所提交的委员会等项目也分别在数据库的记录中给予了反映。其中关于参议院和众议院的听证会情报则包括会议讨论的议题和听证会的细节。

由于考虑到检索者所输入的大部分词条都不是正规的形式,所以 CRECORD 采用了一个标准的缩略语词表来标引委员会及其分支机构的名称。

CRECORD 数据库很好使用。由于它的结构设计得可使用户能最大限度地利用联机机时,所以它很容易被人们接受。

§ 1.5 FEDREG (美国联邦政府纪事文摘)

由首都服务公司(CSI)编制的另一个数据库是 FEDREG,它是《*Federal Register* (联邦政府纪事)》的联机索引。《*Federal Register*》是一个由政府每天出版的有关联邦管理法规(CFR)最新消息的刊物。CFR 有四个部分:1) 规章制度;2) 提议的规章制度;3) 通告;4) 总统行政文件。它还包括法律公告和由政府各部与咨询委员会等召开的会议或听证会的有关资料。在为了贯彻执行立法和国会的意图而编纂的行政程序和机构规则的 100 多卷文件中,CFR 就占有 50 个题目。

FEDREG 的著录格式比 CRECORD 更复杂。每一篇文献都要标上文献类型,如,规章制度、提议的规章制度、通告、会议、听证会、行政命令、总统公告等等。对于规章制度部分,FEDREG 提供了恰当的 CFR 篇名和部分内容,并与发行出版机构一道提供资料的描述性情报、注释截止时间和公文摘要号码等。著录规章制度部分时,针对每一条规则都给予一个有效数据和一个事先规定好的参考数据。对于听证会,则著录了主持会议和请求召开会议的机构、会议讨论的题目或发布的议事日程通告、会议的地点和有关的数据。行政文件的著录包括文件的编号,而摘录的通告则类似于 CRECORD 中的评论的附加部分。CRECORD 数据库中使用的分类代码,在 FEDREG 中同样适用。

§ 1.6 FEDEX (联邦活动索引)

FEDEX 数据库涉及美国联邦政府的所有活动。其资料来自《*Congressional Record*》、《*Federal Register*》、《*Weekly Compilation of Presidential Documents* (总统文件每周汇编)》、《*Commerce Business Daily* (商务日报)》以及《*Washington Post* (华盛顿邮报)》等,时间范围从 1976 年至现在。自 1979 年开始,其资料来源又扩大到《*F & S Index* (世界公司与产品统计)》。FEDEX 由首都服务公司(CSI)和美国预测公司(Predicasts)共同编制,而且该

库还可提供对政府活动中特殊事件的一些简略而又精彩的评论。但该库中有一部分内容与 CRECORD、FEDREG、CIS、PTS INDEXES(以前的 F&S Index)和 USGCA (美国政府贸易合同裁决索引)等数据库相重叠。因为该库收集了《华盛顿邮报》中所报道的有关政府活动的消息,所以它是获取最新消息与事件的一个相当不错的信息源。

FEDEX 数据库使用了由 Predicasts(美国预测公司)用于其 PTS PROMT (世界工业产品市场与技术概况)与 PTS INDEXES(世界公司与产品统计)数据库中的产品和事件代码系统(见第 10 章)。

§ 1.7 VOTES(国会议案表决事项)

VOTES 数据库记录了国会成员的个人表决事项(不象 CRECORD 那样仅记录总的唱名表决结果)。该文档的记录有如下两种类型:

1) 立法的史料——包括法案的编号、法案的主要名称、标引词、案卷索取号、唱名表决的日期、唱名表决程序的类型、总的类目表以及《Congressional Record》的专业资料。

2) 国会成员(议员)传略——包括姓名、州名、地区名和所在党派名称。有时也包括法案编号、主要法案名称、案卷索取号、唱名表决的日期、唱名程序的类型以及国会成员个人的表决记录等。

该文档也使用了在介绍 CIS 和 ASI 数据库时描述过的主记录和分析记录的格式。在 VOTES 数据库中,主记录包括关于法案和国会成员的普通情报。如:法案的名称和国会成员所在地的州名与他所参加的党派名称等。分析记录是唱名表决及实际表决的统计结果。在立法过程(历史)的记录中,分析记录包括参众两院总的表决统计情况以及议员们的档案记录,还包括每一个国会议员在表决时是投票赞成还是投票反对的情况。

如同 CIS 编制的几个文档一样,凡是含有检索策略中列入的词条的那一部分的文献记录,都可采用标准的 PRINT 指令将它们打印出来。意识到 VOTES 文档的这一功能是特别重要的,我们可以把输入一个国会成员的名字组配一个法案编号的检索方法作为基本的策略,因为这两类检索词仅出现在主记录中,所以除非使用专门的打印方法,一般只有主记录才会被打印。

PRINT COMPLETE 指令可将包括实际的表决情报在内的所有分析记录全部打印出来。

要查找所有国会成员投票赞成或反对的记录,则必须在唱名表决字段中执行一个字符串检索。在建立的字符串检索中,为了获得正确的结果,根据想要的分析记录以子集的方式选用一个检索词是比较重要的。在本章后面所举的立法史料检索实例中就有这种检索技术的一个例子(见图 1.2)。

大约有 100 多个范围很广的主题词被用于索引该数据库,这些在 SDC VOTES 手册中都有详细的说明。法案题名在该库中也是可检的,但这些简短的题名并不是官方正式给予的,而是数据库的提供者根据文献内容所起的。

§ 1.8 USGCA(美国政府贸易合同裁决索引)

USGCA 是一个涉及美国贸易程序的联机数据库。一旦某个美国贸易合同被裁决,那么