

你知道文献技术吗

〔法〕雅克·肖米埃 著
李大年 周树森 译



科学技术文献出版社

37.216
290

你知道文献技术吗?

〔法〕雅克·肖米埃 著
李大年 周树森 译

Z1C447/09



内 容 简 介

本书为法国文献技术专家雅克·肖米埃所著，内容丰富，深入浅出。全书共分六章：第一章文献工作的历史，第二章文献工作系统，第三章文献工作手段，第四章文献语言，第五章文献工作自动化，第六章文献网络。

可供图书及情报工作者参考。

你知道文献技术吗？

〔法〕雅克·肖米埃 著

李大年 周树森 译

科学技术文献出版社出版

中国科学技术情报研究所印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

开本：787×1092¹/₃₂ 印张：3 字数：61千字

1982年6月北京第一版第一次印刷

印数：1—13,700册

科技新书目：24—59

统一书号：17176·327 定价：0.32元

译 者 的 话

本书作者Jacques Chaumier是一位法国文献工作专家，除本书外尚著有《文献信息系统》（1972年）、《数据库》（1976年）、《文献分析》（1977年）、《文献语言》（1978年）等书。本书于1971年出版，至1979年已出到第3版。本书内容丰富，且深入浅出，可作为图书和情报专业的教学参考书。

凡具有中等以上文化水平者，均可阅读。由于译者水平所限，错误之处在所难免，欢迎读者批评指正。

1981年春于北京

目 录

译者的话

第一章 文献工作的历史.....	(1)
第二章 文献工作系统.....	(7)
第一节 文献工作链.....	(7)
第二节 文献中心.....	(14)
第三章 文献工作手段.....	(17)
第一节 情报的贮存.....	(17)
第二节 复印技术.....	(20)
第三节 情报检索.....	(22)
第四章 文献语言.....	(34)
第一节 层次结构语言.....	(36)
第二节 组配结构语言.....	(44)
第三节 人造语言.....	(55)
第五章 文献工作自动化.....	(57)
第一节 管理.....	(57)
第二节 利用索引刊物查找和提供情报.....	(59)
第三节 直接查找与定题提供.....	(65)
第四节 文献系统的软件.....	(77)
第六章 文献网络.....	(80)
第一节 文献合作网络.....	(80)
第二节 情报检索系统.....	(81)

36669

第一章 文献工作的历史

图书馆工作意味着保管资料，文献工作意味着如何使用这些资料。前者在很早以前就已被人们所了解，而后者则是在后来才出现的。

事实上，这是奥特莱 (Paul Otlet) 在十九世纪末所从事的工作，这个时期标志着文献工作的历史及其技术的开端。然而在这之前，一些文摘通报已经问世，如1830年出版的《化学总览》(Chemisches Zentralblatt) 及1885年出版的《工程索引》(Engineering Index)。

文献工作历史上第一个重要的事件就是由奥特莱和拉封丹 (Henri Lafontaine) 于1892年在布鲁塞尔创立了“国际文献目录局”。这两位比利时律师把自己一生中的大部分精力都贡献给文献工作、国际十进分类法和国际合作。奥特莱 (1868—1944) 和拉封丹 (1853—1943) 于1895年又把国际文献目录局改名为“国际书志学学会”，后者按著者和系统编辑一种卡片目录。从1895年至1900年，奥特莱汇集了1700万张卡片，并与杜威 (Melville Dewey) 合作提出国际十进分类法。奥特莱于1910年组织召开了第一届世界文献目录大会。

1912年，第一次出现了以缩微形式存贮情报的缩微平片。华盛顿的国会图书馆是世界上第一家配备有缩微照相机的机构。

自第一次世界大战结束后，人们越来越强调如何利用图书馆的资料。杂志或期刊名目繁多，发展很快。文献工作一词是在1931年才开始正式出现的。国际书志学学会又改名为“国际文献学会”。这不仅仅是名称的简单更换，而是反映出文献工作在世界上所起的新作用。1932年，在法国也创立了第一个文献机构，即“法国文献机构联合会”（UFOD）。

文献工作技术始于1930年左右。在1929年至1933年期间，布利斯（H. G. Bliss）首先对书目分类法进行了研究，而兰加内森（R. S. Ranganathan）则奠定了分类法理论的基础。在美国，McBee公司开始研究边缘穿孔卡片的使用。

当时，用于文献工作的工具已经问世，如由国际知识合作研究所出版的第一种期刊缩略代码。1935年在法国也出版了第一份文献机构索引。奥特莱则于1934年发表了他的文献工作专论。该文在以后若干年内一直是此专业领域内的一本基础著作。

三十年代末，文献工作在世界范围内已初具规模。当时一个重要的事件就是1937年在巴黎举行了第一届世界文献工作大会，并于第二年在海牙以国际文献工作学会为基础成立了国际文献联合会（FID）。该联合会的活动立足于各个国家委员会的工作。作为国际文献联合会成员的法国文献工作委员会是1939年成立的。国际文献联合会在某些方面，尤其是在传播国际十进分类法方面是起了重要作用的。该联合会的一个下属机构负责国际十进分类法的日常更新工作。

与此同时，文献工作技术也得到了更快的发展。例如着手研究如何在文献工作中使用穿孔卡片。这种机器于1940年首先在美国应用。1940年，美国试验成功了第一台胶片检索机。

在法国，文献工作是一步一步组织起来的。1943年出现了第一批有关文献工作的法国标准。1945年由法国文献机构联合会开设了文献工作教育课程。1946年，科多尼埃(Gérard Cordonnier)发展了目视重叠比孔卡的使用。在同一时期，英国的巴顿(W.E.Batten)也在帝国化学工业公司应用了这一方法。但应指出，目视重叠比孔卡的原理早在1915年已由泰勒(H.Taylor)提出，并在美国申请了一项专利。依据同样原理建立的法国 Sphinxo公司的卡片于1923年在法国也申请了专利。

第二次世界大战以后，文献工作及其技术得到了更为广泛、更为迅速的发展。1948年，一台快速检索机投入实际应用。1949年，在联合国教科文组织的赞助下召开了国际科学文献分析大会。会议期间发表了大量的论文和报告，尤其是有关穿孔卡片机应用方面的论文和报告。那时，美国许多文献中心都配置有这种穿孔卡片机。在这些文献档中，应当提一提凯西(R.S.Casey)、佩里(J.W.Perry)、贝里(M.M.Berry)和肯特(A.Kent)于1951年发表的关于穿孔卡片应用的著作。该文标志着文献工作自动化进入了一个新的阶段。1952年，萨曼(J.Samain)论述了缩微胶片检索机(Filmorex)的原理，并在两年后予以实际应用。这种缩微胶片电子检索法获得了某些成功并在很大程度上对法国文献工作自动化的发展做出了贡献。在那时，穆尔斯(Calrin N. Mooers)也发表了他的研究成果，尤其是对 Zatocoding 法的描述。人们是从穆尔斯那里采用“情报检索”这一术语的。

1950年，法国创办了第一个文献工作教育设施，即国家文献工作技术学院(INTD)。该学院隶属于国立美术工

艺学院 (CNAM)。

与穆尔斯的研究成果可以相提并论的，就是陶伯 (Mortimer Taube) 于1953年继巴顿之后提出的“单元词”组配编目法。但是第一个为目视重叠比孔卡提供理论基础的人是巴顿。两年之后，陶伯发表了他的重要著作：《单元词标引法和操作手册》。

佩里和肯特这两位文献工作自动化的先驱，曾于1956年和1957年相继发表了两篇重要著作：《机械化的文献检索》和《文献检索机》。

五十年代末，研究成果极为丰硕。1958年，西特伦 (J. Citron) 和奥尔曼 (H. Ohlman) 介绍了“轮排索引”的思想。同一年，IBM公司的罗恩 (H. Lubn) 阐明了“定题情报提供”系统的概念。

有关这方面的国际会议也日益频繁并吸引了越来越多的专家。1957年在伦敦举行了文献检索分类法国际大会。1959年在克利夫兰召开了国际文献和机器翻译通用语言大会。

可是，当时法国的文献工作并未达到自动化程度，只是到了1958年，巴斯德研究所和佩希奈公司才对如何使用穿孔卡片机进行了初步研究。这些尝试有的以失败告终，另一些则中途夭折。一直到了六十年代初才有了新的发展。

从1960年开始，文献数量急剧增加，人们开始谈论起情报爆炸的现象。根据联合国教科文组织统计，十九世纪初期仅有上百种科学期刊，1850年已达一千种，1900年是一万种，而今天已超过五万种，其中法国期刊有两千多种。最近十五年来，文献数目按指数增长。1958年，科技文献的世界年产量已超过100万篇，约1300万页。1961年达到200万篇。

1967年，估计为350万篇左右。1970年，已超过400万篇。

1962年，法国图书馆管理处出版了一本图书馆和文献中心目录，登录在该目录中的文献中心或图书馆有2352个，尚不包括为数众多的属于实验室、企业或联合会的中心。

1973年，法国成立了国家科学技术情报局（BNIST），旨在协调全国文献工作。

随着文献数量的急剧增加，文献工作自动化方面的成果也增多了。穿孔卡片机已由电子计算机代替。在法国，Saint-Gobain公司首先将电子计算机用于文献工作（1960年）。1968年，根据国家科研中心科学技术情报研究组编制的一份统计，应用的成果已达23项。1971年，根据法国文献工作者和职业图书馆馆员协会（ADBS）的一项调查，采用计算机进行文献情报加工的已有36个中心。1976年，由该协会进行的又一次调查表明，使用计算机技术的中心已达120个。

随着具有大容量存贮和随机存取功能的第三代计算机的问世，计算机在文献检索中发挥了更大的作用。目前在文献服务部门中，通常使用的现代化设备包括联机系统、计算机网络和专用终端。专门用于文献情报处理的欧洲网络发展了计算机在文献工作中的应用。这就是由欧洲经济共同体部长理事会决定建立的EURONET网络。该网络于1979年投入使用。

除了图书以外，文献还包括期刊文章、会议录和研究报告。同样，除了纸张以外，新的情报载体还有胶卷、缩微平片和磁带等。

文献工作已成为一门专业，具有自己本身的技术和特有

的方法。这一学科求助于语言学、数学和计算技术等一些科学。正如一些专门杂志所阐明的那样，文献工作具有自己的学问。

第二章 文献工作系统

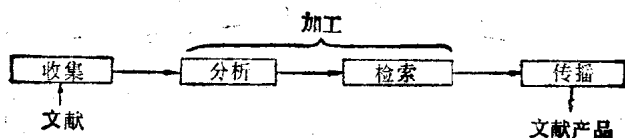
文献工作系统的类型可根据各个中心所承担的任务或所确定的使命而定。

第一节 文献工作链

文献工作的各个环节是按一个既定顺序首尾相接的，并且不论是哪一种文献工作系统，这种顺序都是大同小异的。这些环节构成所谓的“文献工作链”。

温伯格（Weinberg）对这条链作了如下描述：情报工作的过程包括一些分开的阶段或“基本的操作”：生产、记录、整理、编目、保管和传播，并由用户进行查找和使用。既然各个阶段是首尾衔接、互为依赖的，那么情报工作的这整个过程就好比一条链，人们也称其为“情报传送带”。

这条链的各个环节可由下列示意图予以表示，该示意图还显示出整个文献工作系统活动的线性特点：



一、收集

收集是文献工作链的第一个环节。收集文献是任何文献工作系统的首要功能，有时甚至是一个中心仅有的功能并代表着该中心的最终目标。为了有效地进行文献的收集工作，第一步措施就是在一个（至少）中心点设置法定的保管机构。

遗憾的是，由于“地下情报”的出现，即那些未通过正常出版渠道而发表的文献，如实验室或机构团体的研究报告、论文、预印本（正式出版前所散发的印数有限的文献，而实际上并未予以出版）、会议或研讨会论文等，使收集工作成为一项费时、复杂和昂贵的作业。

人们并未采取什么措施来协调一次文献的生产和传播，包括一般出版工作（图书和期刊文章）。然而在这里要提一提某些国家的一些机构曾为整理和传播这类难得文献进行过努力，如美国的国家科学技术情报服务中心（NTIS）。

二、文献情报处理

“情报处理”是指对文献进行加工、整理、存贮和再现的整个作业程序。作为文献工作系统基础的文献情报处理包括两个主要阶段：分析和检索。

1. 文献分析：这是指用一种异于原始形式的方式来表示一篇文献内容的一个或一组工序，以便于下一阶段的查考或标记。不论是什么样的文献，都不能按其原来的形式记录到一个文献系统中去，这是由文献的篇幅、复杂性和其表述方法所造成的，必须按一个通用的模式并根据既定系统的要求对文献予以加工，才能进行存贮和检索作业。文献分析是一项脑力劳动，其复杂程度要视所要求的加工和精细程度

而定。这项作业的产品就是二次文献（这是与被称为一次文献的情报相对而言）。产品是多种多样的，这取决于所使用的分析方法。

第一种最熟悉的方法是编辑文摘（有时也称为摘要）。文摘是一篇文献所含信息的压缩形式，其长短要根据分析水平，文献的篇幅或所使用的文献系统而定。文摘可以是几行文字（报道性文摘）或者十几行、几十行（指示性文摘），甚至于几页文字（分析性文摘）。人们也可以根据其来源予以区分：由著作者本人编写的作者文摘和由专门文摘员编写的文摘。尽管人们对作者文摘缺乏点客观性予以某些批评，但这种文摘形式还是发展很快的。在某些技术领域内，将近60%的文献都附有作者文摘。一些刊物或机构要求著者提供一份文摘。国际标准化组织的一项《建议》专门论及作者文摘的编辑，并为如何编写作者文摘出版了一本手册。对于作者文摘，温伯格在他的报告中是这样讲的：“著者本人更应当意识到自己对追溯检索的责任……。他们应当选用确切而含意丰富的术语，以便对读者给予积极的帮助……。任何人都不能像著者本人那样经济地对其文章进行分析和做出文摘”。

最常见的文献分析方法是标引。标引就是使用一种受控语言（相对于自然语言）中的术语，以某种程度的压缩方式来表征一篇文献的全部内容。在整个文献工作链中，这个环节是必不可少的，它确定了供存贮与检索工作使用的全部数据或判据。标引时，关键在于如何抽出一篇文献的所有概念并借助于关键词、叙词或分类标志予以表达。标引的粗细程度要根据所使用语言的复杂性和该系统所要求的目 的 而 定

(用于定题情报提供或追溯检索)。标引词一般不同于抽词。在标引时,从文献中抽出的概念或表达这些概念的词常常由所用语言中的另一个词代替(例如复制技术代替了复制)。另外还使用一种分类标志来进行标引。在抽词时,仅分析文章中那些看来是表达该文主要概念的词,而不进行任何语义转换。在标引时,则进行词法上的加工,以便缩减表达同一个概念的各种形式(例如仅使用单数和阳性)。

当文献输入某一特定系统时,还可以对文献进行进一步分析。这主要是指对全文,或者题目进行翻译(例如准备编制多语言对照的文献题录索引)。

2. 文献检索:为了建立文献档,就需要将经过分析的文献存贮起来。文献的检索就是在这个文献资料档上进行的,这项操作还称为选择作业。

(1) 文献可以按照以下两种组织形式来存贮,即按文献组织和按概念组织。

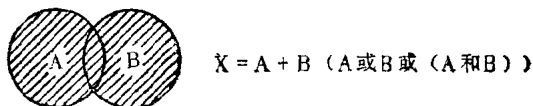
如果系统是按文献组织的,一篇文献的全部特征(分析时抽取出来的特征)要标注在按文献建立的卡片上。记录单位就是一篇含有全部特征(关键词、分类号等)的文献,而且有多少篇文献就有多少个记录,这种组织形式用于穿孔卡片或顺排资料档。另外还可以在一张文献卡片上记录一篇文献的一个特征的形式来组织文献。在这种情况下,一篇文献的记录数就等于标引时从该篇文献中提取的特征数。这种记录方法就是按内容分类的普通资料档方法,对这个资料档来讲,每一篇文献的标引词数目与记录数目相等。

如果系统按概念来组织,全部文献都贮存在特征卡片(或概念卡片)上。这样,记录单元就是一张特征卡片,在

特征卡片上记录含有这一特征的全部文献。存贮的记录数目与本系统所使用的分析语言的词数相等。这种组织形式就是单元词资料档形式，也称目视检索档或倒排资料档。

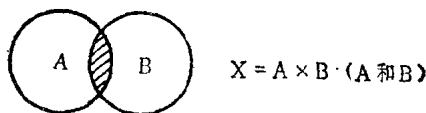
(2) 对存贮的文献进行检索时，可利用逻辑代数（布尔代数）和在文献集之间加逻辑关系的方法来进行，每个集均由一个分析特征来确定。在文献检索中使用的主要逻辑关系是逻辑或、逻辑与和逻辑非。这些关系由逻辑算子“或”、“与”、“非”来表示。对于一篇由特征A和B表示的文献，A、B逻辑关系可以用一般的“欧拉圈”表示，即每一集合用一个圈表示。在该圈中，根据所实施的运算，各集相应的部分就是这些圈中加晕线的部分。

各类的“并”能回答这样的问题：从存贮库中选出由A或由B，或者由A和B所表示的全部文献（逻辑“并”使用“或”算子）。根据欧拉圈的表示，人们可以得到下图：



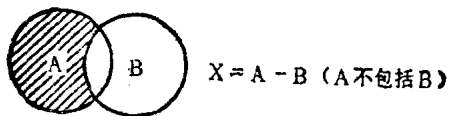
例：查找有关“文献工作（A）”或某些“图书馆（B）”或者有关“文献工作”和某些“图书馆”方面的文献。

利用各类的“交”可以找出由特征A与B表示的所有文献（逻辑“交”使用“与”算子）。简图如下：



例：查找有关“文献工作（A）”与某些“计算机（B）”的文献。

利用逻辑“非”可以找出除了特征B以外的所有由A表示的文献（“非”算子）。这一逻辑非由下图表示：



例：查找有关“复制技术”方面的文献（A），但有关“缩微胶片”的文献（B）除外。

其他逻辑算子：小于、大于和等于在文献系统中很少使用。检索时，可以利用逻辑算子组成各种提问式，以便得到比较准确的检索结果。在文献检索阶段，人们常常将提问式称为“检索策略”。如果一个提问的所有术语用“或”算子连接，则输出的篇数就增加；如果用“与”算子连接，则输出的篇数就减少。

检索原理是基于提问的全部特征和存贮文献的全部特征之间的联想相似性，所获得的答案永远不是十全十美的答案。可用误检率、漏检率和查准率来说明文献系统的性能。

误检就是在输出结果中答非所问的文献所占的比例，主要来自不恰当的组配或术语在概念上的混淆。对一个提问式，如“情报技术的历史”而言，三个特征的“交”也将会反映另一种组合，这个组合能对下述问题：“关于古代技术的情报”作出响应。论及这后一个主题的所有文献就是系统对所提问题的误检。

与这个概念相对立的就是漏检。漏检就是在存贮器中的