

日本科学技术厅 编

科技情报工作
现状和展望丛书

化学情报

科学技术文献出版社

科技情报工作
现状和展望丛书

化 学 情 报

〔日〕科学技术厅 编

张翠兰 译

周淑贤 乔 杰 校

科学技术文献出版社

1987

内 容 简 介

化学是关系各学科领域的一门科学，是重点研究物质的学问，因此人们很早就把化学作为情报系统的一个典型领域进行研究。化学情报内容包括文献情报、数值情报、化学结构的标记和处理，以及情报的收集、处理、流通过程的技术问题和实现计算机化等一系列问题。本书对其动向、问题和展望予以评论和介绍，并在化学情报各种问题中，提出人们最关心的热门议题，以帮助读者，特别是从事教学、科研、生产和情报工作人员，了解、掌握并利用国际上化学情报科研成果。

本书可供从事教学、科研、生产和情报工作人员阅读参考。

科学技术厅 编

化 学 情 报

大藏省印刷局

昭和54年5月

化 学 情 报

〔日〕科学技术厅 编

张翠兰 译

科学技术文献出版社出版

北 京 印 刷 二 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 11.25印张 240千字

1987年5月北京第一版第一次印刷

印数：1—3,500册

科技新书目：142—57

统一书号：15176·765 定价：2.30元

出版说明

《科技情报工作现状和展望丛书》是日本科学技术厅主编的，内容包括：化学情报；医学情报；科技情报的机械化和自动化；农林水产科技情报；冶金情报。我社翻译出版的这套丛书，先后次序与原书有所变更。现已出版的有：《医学情报》、《农林水产科技情报》、《科技情报的机械化和自动化》，这次出版《化学情报》，其他将陆续出版。

1986年10月

译 者 的 话

随着科学技术的迅速发展,一些技术先进的国家对于大量的化学情报如文献资料、化工技术经济情报以及表征物质结构的各种谱图等,均可利用计算机文件或数据库的形式进行存贮、管理和检索。

《化学情报》这本书较全面地阐述了化学文献情报、数值情报、化学结构的表现和处理、术语解释和有关化学数据库等方面的现状、问题和展望。本书于1979年5月出版,因受时间限制,故所收集的内容多为七十年代水平。进入八十年代,国外计算机技术发展很快,有些新的成就尚未汇集在内。

近年来,我国利用计算机进行化学情报的处理与检索方兴未艾,但与国外先进水平相比尚有差距。目前国内尚无全面介绍化学情报的书籍,因此尽管本书出版较早,其中有些问题从日本国情考虑多些,但其所阐述的利用计算机进行各种化学情报及数值情报的存贮、管理和检索等基本原理,对我国从事这方面工作的科技人员、科研及生产单位还是有参考价值的。翻译出版这本书希望能对促进我国化学情报服务工作的现代化起到一点作用。

由于时间仓促,水平有限,难免有不妥甚至错误之处,敬请读者指正。

译 者

1985年9月14日

· II ·

前 言

随着科学技术工作的迅速发展，科学技术情报的数量在显著增加，形式亦向多样化演变，以致现在日益难以综合性理解和掌握科学技术各领域的工作动向。

因此，科学技术厅从1972年起对综合性科学技术领域着手进行综合评论工作，就其发展的进程、现状、问题与未来展望等加以归纳，编辑对行政、经营和研究管理等方面有用的资料，以促进科学技术情报的广泛流通。

此种综合评论，过去是按海洋科学技术、材料科学技术、电子通信、生命科学以及系统工程，依次从各自领域中提出主题，编写成丛书形式出版的。

鉴于近来人们对科学技术情报工作重要性的认识日渐提高，我们在1977年从“科学技术情报工作”领域提出“化学情报”这一题目，将其作为综合评论的对象，委托社团法人化学情报协会编辑出版这本书。

本书是科学技术情报工作丛书的第一卷，是对文献情报、数值情报、化学结构的标记和处理等方面就其动向、问题和未来展望等进行评论的著作。它可以帮助读者加深理解化学情报在新的化学创造上正在起着怎样的作用。

期望读者将本书与其它领域的综合评论刊物参照使用，同时如蒙对本书坦率提出宝贵意见，则不胜荣幸。

最后，谨向对本评论工作给以协助的各位有关人士致以谢意。

科学技术厅振兴局局长

山 口 和 男

1979年3月

序

情报处理和利用上的机械化，在日本亦将会逐渐普及。大量情报的存储和高效率检索系统是一门综合性科学技术，只有把在语言和符号等信息标记的基础研究领域工作的电子计算机技术、电子和通信工程学乃至特定科学和工程学的专门研究人员集中起来组成一个完整的体系，有机地发挥其功能，才能取得丰硕成果。从这种意义上讲，科学技术情报是一门庞大的系统工程学。

在情报量大这一点上，化学情报与医学情报并列构成科学技术情报之双壁；在需求情报的人数上，化学情报恐怕也是最大领域吧。化学情报作为使用机械处理的对象来看时，具有的有利因素就是化学作为精细科学已完全体系化，是以物质为中心的科学。因为仅是量大却杂乱无章的信息是不适合电子计算机处理的。

从这样的理由出发，化学在情报处理和流通的机械化方面，不仅是处理技术问题，同样在建立国际性合作支援体制上，十五年来一直处于领先地位。现在对情报的收集、处理和利用正以世界规模进行着，无论技术方面或利用率上均处于一种日新月异的形势。

进步快与作为一种系统经常保持平衡向前发展并非同义。不断反复探索，处理存储技术的开发工作先行，而标记方法的确立工作落后；或者使用技术未能充分开发，致使有

潜力而无法施展利用，这些都是出现的暂时不平衡。另外存在的问题，就是象文献分析 (document analysis) 那种依靠人的智能工作的部分尚未实现机械化。

本书是从正在取得这种迅速进步的系统的化学情报各方面中选取了一般认为当前最热门课题的部分，作为充分了解现状以探索未来前进之路的里程碑而规划编写的。因此，对系统中许多重要的部分，在编写中未能涉及到。比如对一次情报刊物的编辑、学术用语的规范、文献分析的技术以及高次情报利用等题材均割爱未收。

目 录

第1章	绪 论	(1)
第2章	文献情报	(10)
2.1	总 论	(10)
2.2	情报的流通系统	(15)
2.3	情报的提供	(26)
2.4	二次情报制作过程	(34)
2.5	日语的处理	(60)
2.6	专利情报	(65)
第3章	数值情报	(75)
3.1	数值数据的评价、收集和检索工作的现状	(75)
3.2	关于分子和结晶结构的数据及相关数据	(80)
3.3	热物性和高压物性数据	(95)
3.4	电化学数据	(108)
3.5	光谱数据	(125)
第4章	化学结构的表示和处理	(141)
4.1	化学结构的线性标记法	(141)
4.2	结构情报的处理	(174)
4.3	化学结构的检索	(221)
4.4	蛋白质结构数据的检索	(233)
第5章	存储、检索和输出的技术问题	(247)

第 6 章 其它有关问题.....	(290)
结束语.....	(300)
附录 1 有关化学的数据库一览表.....	(302)
附录 2 术语解释.....	(332)

第1章 绪 论

1.1 所谓化学情报论^[1]

本书主要是对化学领域中情报的处理、流通和利用的现状进行概述。在后面分章叙述之前，想先就全面性事项，对情报论的各个侧面及其相互关系予以记述。

在科学技术和医学领域，一般认为基础的研究，在于认识自然界，重点应放在丰富和深入掌握知识上；应用的研究，则是关于如何把这些知识转变为有用的技术和制品。为了学术得以发展，我们应当尽快到达前人所开拓的新领域的最尖端，使新领域的尖端科学技术不断前进。情报的取得、流通和利用是为达到此目的所必需的手段。这里所谓的“取得”是狭义的意思，是要把大地上不论何处已经成为人类之物的知识取为己有。从广义上讲，对某种物质进行前所未有的观察和测定而获得数据，也是情报的取得。这纯属科学技术研究问题，故这里不去考虑。

情报论尚未作为一门学术体系予以确立。过去分散于诸多学术领域的研究成果和研究手段，现正在从新的角度予以重新估价，使它们相互关联起来，以“情报论”、“情报科学”、“情报工程学”等名称加以重组，努力做到体系化。同时，开始对其不足之处进行研究。

情报论有与其它科学技术和医学等学术不同的一大特

点,即后者的重点在于去认识自然,从而去利用自然,而情报论则相反,它是一种旨在为了推进人的智能工作的一门技术。化学情报论并非与有机化学、无机化学等并列的化学中的一个专门范围,而是探求为了向化学及其边缘领域的研究人员和技术人员提供更良好的传递方法,把人类的知识作为共同财富传给后代的一种手段的学问。

化学情报论就是从上面这样一个目的出发,处理怎样表示研究人员所得到的知识,怎样将其存储起来,又怎样取出所存储的内容等各种问题。如果解释得广泛一点,会涉及到学术用语、单位名称和符号、物质命名法、物质结构标记法、情报分析法、索引法、记录法、通信技术以及检索法等。

1.2 化学情报的流通^[2]

化学情报从其形态分类,可分为下面两种。

a. 文献情报是揭示所要检索的知识所在的情报。一般是关于论文等的题目、作者、期刊名称、卷号、年份、页码(所谓著录事项)的内容。

b. 学术数据情报(数值和图形情报)是关于具体知识本身的情报。

例如,想了解水的密度时,告诉你可以解答该问题的学术杂志的某一页上刊登的论文中有所记载的是a,具体地告诉你在0℃为0.99984的是b。而想知道关于大气中氮的氧化物的检测方法,已经得出了什么样的研究结果,a可满足要求,b则无能为力。

情报的生产者是研究人员,最终利用者还是研究人员。

但这一流通是经过一个非常复杂的过程，流通过程亦各种各样。首先就a的文献情报研究一下情报的流通。

图1.1是十九世纪以来，经过大约七十年时间，自然形成的情报流通图。研究人员撰写的论文以学术协会发行的学术杂志的形式印刷出版，称为一次情报。一次情报直接送给研究人员或通过图书馆为研究人员所利用。象最近化学情报一年达40万件以上。在这种情况下，开始进行某种研究，完全单从一次情报中找出有关文献是不可能的，所以一般都要借助于文摘刊物等的二次情报。二次情报是把各种各样的一次情报加以整理（分析），以可以检索的形式向研究人员提供的一种方式。没有二次情报，不可能取得有组织的情报。目前作为化学二次情报，规模最大的是美国《化学文摘》（Chemical Abstracts）。另外，为专门学科编印出版的文摘刊物（专利类、药学、金属等）亦可广泛利用。从六十年代后期开始，各国都在大力开展利用计算机处理情报工作，结果如图1.2的下半部分，尤其是二次情报机构的工作发生了很大的变革。即二次情报机构不仅以印刷品的形式提供情

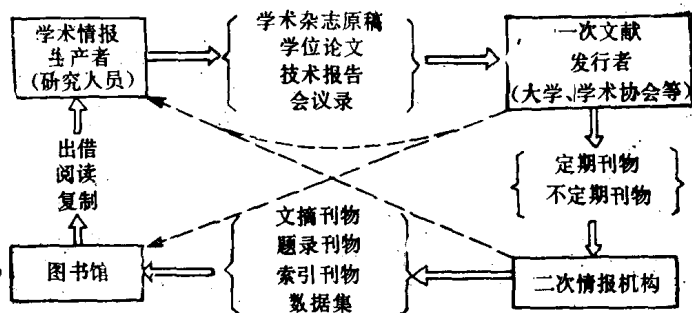


图1.1 古典式学术情报之流通

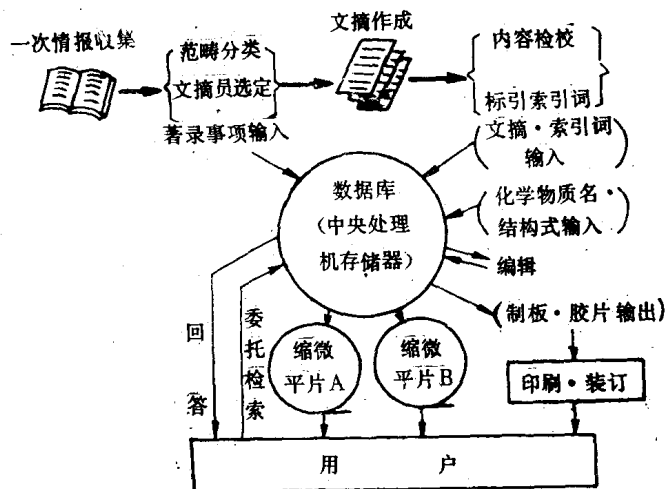


图1.2 二次情报机构作业图

报，而且通过计算机处理成为可能的一种用磁带提供情报的方式，正在日益发展中。图1.2所示，为在二次情报机构已实现机械化的工作形式之一例。这一工作程序包含了化学情报论各主要问题的几乎所有方面，拟在下面对其略作较详细的阐述。

1.3 情报分析^[3]

理想姑且不管，而现实上情报尤其一次情报是以各种不统一的形式出版的。如不从统一的角度加以分析、整理和加工成便于利用的形式，则不能作成可检索的二次情报。对计算机情报系统，如图1.2中输入电子计算机前的阶段，就是这

一部分的工作。输入项目大致区分为以下三种：

- a. 著录事项
- b. 文 摘
- c. 索引词

仅以著录事项来看，就可以理解到按统一的原则输入的重要性。比如杂志名称的表示法任文摘员随意决定，一文摘员写成 J. Am. Chem. Soc., 另一文摘员写成 J. A. C. S., 这样在以后就难能有恰当的检索。另外，还存在一种叫做 J. Am. Ceramic Soc 的杂志，所以情况就更加复杂。

文摘是日常情报处理工作中最有学问的部分。判断原作中哪部分重要，以后需要检索，确实非常困难，这是现行系统中最大的薄弱环节。要作出完好的文摘，就得具有对原文所处理的主题的高度的专门知识，而且对数据库的结构和检索手法亦应了解。对这方面的人才培养，日本尤其落后。原文中含有重要情报而文摘中未予记载，无论后面处理系统如何优佳，除遇偶然机会好不容易重新看到原文外，是全然无法获取的。

索引词的选定有两种方式。一种是在原文作者所用的词汇中抽选出认为对检索原文内容所必要的词，称此为自由词（自然语言）索引，《化学文摘》每期后面的关键词索引（Keyword Index）即属于自由词索引。另一种方式是准备一定的词汇表，将作者用的词汇转换成词汇表中收录的词，然后作为索引词。用于该目的的词汇表叫做主题词表。《化学文摘》的卷末索引大体相当于此。当然，也有属于此二者之间的词汇。主题词表的编制和维护是情报论的重要部门。

索引词中有概念词和物质名。与概念词主题词表相应的

是对物质名的命名法。在国际上得到承认和使用的化学物质的命名法，是国际理论化学和应用化学联合会（IUPAC）命名法。但从计算机处理的角度看，该命名法有不足之处，现正在进行改进。另外，如何将化学结构变换成计算机阅读形式，也是目前正在进行的重大研究和开发课题。

1.4 数据库

数据库一词有几个方面的意思，这里是指最一般的“将各种信息元素按一定方式统一为计算机阅读形式的数据库”而言。将数值、字符、符号变换成磁脉冲序列，记录到电子计算机用磁带上。将经过分析的信息分解为许多元素（作者姓名、标题、刊名、卷、期、页、年月、作者地址、国别、语种、文摘编号、每个化学物质名等），对之各赋予一指定的固定专用编号，按一定顺序排列，与控制代码一同输入计算机。这些信息元素对计算机的存取是通过在该计算机的操作系统下工作的软件进行的。这种操作系统改进的方向是数据库的作成既可以在情报分析全部终了之后进行，也可以在分析过程中进行，当分析终了时数据库也已作成。

一旦数据库建成，便可按软件的设计依次建立各种子数据库，也可以象按成衣样式定做服装那样根据研究人员要求提供专用子数据库。美国化学文摘服务社把子数据库称作文档。还可用计算机制版，所以制作特别定做的印刷品也很容易。

把每年按40万件增加的文献信息和已知的450万化合物（每年增加30万）存储到数据库并要加以维护，这种系统是