

(美)尤金·加菲尔德 著
侯汉清 等 译

引文索引法的 理论及应用

北京图书馆出版社



引文索引法的理论及应用

(美)尤金·加菲尔德 著
侯汉清 陆宝树 马张华 译
刘 煜 侯汉清 校

北京图书馆出版社

图书在版编目(CIP)数据

引文索引法的理论及应用/(美)尤金·加菲尔德著;侯汉清等译. —北京:北京图书馆出版社,2004.8

ISBN 7-5013-2438-7

I. 引… II. ①尤…②侯… III. 引文索引—研究 IV. G353.21

中国版本图书馆 CIP 数据核字(2004)第 039448 号

书名 引文索引法的理论及应用

著者 (美)尤金·加菲尔德 著 侯汉清等 译

出版 北京图书馆出版社 (100034 北京西城区文津街7号)

发行 010-66139745 66175620 66126153

66174391(传真) 66126156(门市部)

E-mail cbs@nlc.gov.cn(投稿) btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com

经销 新华书店

印刷 北京市广内印刷厂

开本 787×1092 毫米 1/16

印张 16.5

字数 350(千字)

版次 2004年8月第1版 2005年5月第2次印刷

书号 ISBN 7-5013-2438-7/G·575

定价 35.00 元

中文版序言

尤金·加菲尔德

1980年我首次访问中国时,《引文索引法的理论及其在科学、技术和人文科学的应用》*刚刚问世不到一年。1983年此书第一版销售告罄,由美国科技信息研究所(Institute of Scientific Information,简称ISI)重印。在我的众多出版物中,这本著作的被引频次无疑是最高的,虽然我尚未做全面的分析以确定被引频次高的原因。许多著者是把此书当作引文索引法或引文分析的符号而予以引用的。有些著者则是引用此书的某些章节,例如在期刊评价中使用引文数据或文献计量学的某些方面。

1980年在中国当时有很少的迹象表明,引文索引法已为人知晓。但是几年后,引文索引法引起了中国学者、编辑以及科研管理者越来越大的兴趣。《科学引文索引》(Science Citation Index®,简称SCI®)印刷版和1980年以后的光盘版,在中国高校图书馆学和情报学系中越来越多地得到使用和讲授。随着互联网和ISI的“科学网”(Web of Science)的出现,SCI在文献检索中独有的链接功能逐渐地显露出来。SCI作为科研和期刊评价工具的功用也明显地表现出来。

尽管SCI在美国及其他地方得到了广泛的普及,但是,相关的用户教育工作仍是一项永无止境的任務。ISI和大多数其他文摘索引机构一直在向图书馆员和科研人员开展大规模的用户教育项目,以推进其产品的普及和正确的使用。然而,短短几个小时的讲座或短暂的研讨会所能做的事情,往往是有限的。由一些已经掌握SCI的科学家担任辅导教师、培训他们的学生,去解决每天实验室中遇到的研究问题,是最好不过的了。《引文索引法的理论及其在科学、技术和人文科学的应用》一书可以向科研人员提供有关SCI内容和义理的全面了解。

在中国不少知晓用户教育需求的高等院校教师中,经过纽约长岛大学储荷婷教授的介绍,我结识了南京农业大学侯汉清教授。他提议,由他及他的一些同事把《引文索引法的理论及其在科学、技术和人文科学的应用》(1979年版)翻译成中文,尽管此书出版距今已有20多年,并可以在网上免费获取,而且大多数研究者都用英语阅读科技文献。我同意了他的建议,因为引文索引法和引文分析的基本原理还没有改变。事实上在SCI问世近40年之后,基于引文的检索已经成为一种惯用的检索方法。不管用户是从SCI的一个题名关键词开始查找,还是利用它的其他索引查找,追踪一条被引的参考文献展开搜索,现在已经几乎成为一种常规的做法。当然,那些熟悉SCI的用户会从一篇已知的论文或一本图书出发,并扩展其检索。

2002年我访问中国,先后到过北京、合肥、上海、成都和香港等地。我在上海停留期间,会见了侯汉清教授。我们在一起讨论了这篇中文序言的主题。为了赶上这本书出版的截止期,我们一致认为,要对过去25年中已经发表的众多相关文献予以评述,是不大可能的。前些年

* 简译为《引文索引法的理论及其应用》——译者注。

已经发表了多篇重要的述评文章。而且述评的主题仍在不断地更新。中国的期刊编辑和科研人员似乎更加关注于期刊影响因子这一论题的讨论。最近我发表的关于这些讨论的论文,已经全都放在我的个人免费网站上。

最后我还要指出,有关此书第七章阐述的引文分析用于科学史的图示这一论题的讨论,近年来十分活跃。我们已经开发出将 SCI 的相关数据转换为科学史图的软件,详见美国情报学会 2002 年年会的报道(http://garfield.library.upenn.edu/paper/asis2002_presentation.html)。这些方法与“文献管理者”(Reference Manager)之类工具的结合使用,将会使得科研人员管理个人文献的工作变得更加有效。引文索引法和引文分析的最终目的就在于促进获取和发现更多的创造性情报。

英文版序言

罗伯特·默顿*

我受尤金·加菲尔德之托将此书介绍给读者,并愿借此机会谈谈我对此书普遍意义的一些看法。此书系统地介绍他对科学发展的特殊贡献——第一部科学文献多学科引文索引的历史、方法和含义。虽然我是尤金的一位亲密朋友,但是这并不能解释他请我做序的原因。尤金有许多好友,他们对情报学和引文分析法的造诣远比我深。我认为,尤金做此选择的意义正是在于他矢志要成为一名科学社会学家,而不是成为一名情报学家。尤金除了具有丰富的想像力、独特的判断力和发明、生产与开发有效文献检索工具(查阅文献是科学活动中一项似乎平常、但是却十分重要的工作)所需要的旺盛精力之外,他最突出的特点之一是他对潜藏于科学实践中的社会、文化和认识结构的直观认识较为深刻。这使他从一开始就认为,他发明的SCI将发展成为科学史研究和科学社会学研究强有力的工具。我认为,他意在使此书的主题尽量涉及这个较广阔的领域,而不是仅仅局限于情报学这个虽然重要但比较狭窄的范围。我欣然接受他的委托,但是也有一点人们可以理解的不安心情。

此书要是早出版就好了,它最直接的指导意义在于它可以帮助读者了解引文索引的概念和作用。作为一种文献检索工具来说,利用引文索引查阅文献比起利用以往的索引不但查得全,而且也简便得多。因此,本书显然满足了人们的这种要求,即系统描述这种旨在促进科学活动的大型书目工具的概念和使用方法。因为本书立论清晰,论据充分,这正是人们对这种书目工具的创制者的希望和要求。

本书对于满足用户掌握作为一种实用的书目工具检索文献情报的引文索引的需要,将起到重要的作用,但是我对此书的主要兴趣则不在于此。这种看法的形成当然是一个社会学家的观察结果,因为社会学家的主要兴趣是致力于研究:相互作用的科学认识结构和科学社会结构是如何影响科学家思维和行为的。也许有人会说这是一种带倾向性的看法。从上述观察来看,实际上引文索引本身就为发明和发现的过程提供了一次案例研究。即使从本书的大致内容来看,它也始终没有脱离社会学方面的内容。

按照历史发展过程系统地利用科学论文和图书中的参考文献和引文,能够推测出潜在的科学的社会结构与文化结构。我们可以先讨论这个问题。看来,这似乎是科学界中的一个自相矛盾的问题。因为科学家能使其他人任意利用其知识财富的范围越大,其财富得到公认的把握性就越大。科学知识属于大众而不是属于私人,所以科学家只有通过发表他们的著作才能有所贡献;也只有当其著作成为大众的科学财富的一部分时,他们才能真正将其称为他们的著作,因为这种著作的资料来源必须得到同行们的承认。一个成就卓著的科学家的最大愿望莫过于从事某种被同行大量利用和十分尊重的工作,因为他们最有资格评估它的价值。一般

* 罗伯特·默顿(Robert K. Merton)美国科学社会学的创始人,哥伦比亚大学教授——译者注。

说来,为人尊重的科学工作的标准是他人可以借鉴其来推动他们未来研究工作的进程。

所有这一切都由于科学奖赏制度而得到了加强。由于得到有资格的同行承认是外部奖赏的一种基本形式(所有其他形式的外部奖励均源于此),加之要在著作公布于世时才给予报酬,所以这种历史形成的奖赏制度是对公开出版著作而无直接经济报酬的一种惯例性鼓励。同行承认通常是与在某一方面首次发表著作相一致的,其后所发表的类似的著作大概会成为多余。但是,由于科学的认识结构可以帮助人们做出独立的多种发现(功能上等同但不是完全相同的发现),所以这种社会的和认识的复杂性会使科学家注意捷足先登和通过迅速发表著作取得其自认为有效发现的优先权。

侵犯科学领域中的知识产权问题只有通过公开发表的著作才能被充分证实,出现这类问题则与科学家承认利用过他人著作的道德水平与认识要求有关。因此,引文和参考文献只在共同的认识和道德范围内产生影响。从认识方面来说,它们旨在提供知识的历史的内在联系和指导新著作的读者发现他们曾借鉴的资料出处;从道德方面来说,他们是通过公开承认利用过别人著作这一惟一形式来偿还知识上的债务。这种偿还绝非是一种无足轻重的要求,从被确认为不同程度地剽窃过他人知识,即现在称为的抄袭行为触犯上述准则时将受到道德上的谴责,有时甚至被法律制裁这一点,就会明白一切后果。马克思针对违犯准则的行为深刻地指出,抄袭是对私有产权的一种露骨的和极其可恶的犯罪行为。他不但将这种看法写进了《资本论》第一版的前言中,而且在整个这部革命性的著作中对其进行了进一步的无情抨击。

对于要得出其引文索引概念的尤金·加菲尔德来说,当然无须对上述这种复杂的交流——知识产权和奖赏制度进行鉴定。他只需确定通过引文索引法,他所建立的系统能够提供各种元素,以系统地反映科学家的著作之间的联系,因为这种联系既可以用于检索文献,又可以用于探讨科学中的认识层面与社会层面的相互关系。本书大约用了一半的篇幅介绍后一种发展的引文分析方法。

尤金·加菲尔德在描述其引文索引法的产生和发展时,报道了在其发明过程中发现的其他有代表性的模式。详细的过程及方法的研究有时是以小型试验的形式进行的,其目的在于检验不断扩大不同应用的可行性。自我批评和外界批评推动了寻找差错和纠正差错的研究。这种有组织的自我怀疑活动迫使我们去进一步掌握科学交流与奖赏制度及其认识发展之间的相互作用的社会学知识。由于社会学家 Norman Kaplan^[1] 10 多年前首先发现这个问题和此后引文分析^[2] 越来越多地应用于社会学中,所以如果引文分析是为了进一步了解科学的社会结构与认识结构的形成和科学实践是如何进行的,我们就需要更多地了解参考文献与引文能够说明哪些问题和不能说明哪些问题。

引证行为的某些模式似乎会限制引文计量在考证某种学说源流中的应用。其中之一被描述为“合并消灭模式”,即通过将某种学说、方法或发现并入新接受的知识中使其失去资料的出处^[3]。在这一假设的过程中,论文或著作中公开引证其参考过的原著的次数减少。因为这种知识的用户和后来的传播者对原著的内容已经了如指掌,所以就会假设读者也是如此的,由于他们不愿损害其读者的面子,所以再也不会提及这些原始资料的出处。加之我们中的许多人都有这样一种倾向,即总是认为某种重要的学说或公式是那些将其介绍给我们的著者提出来的,所以完全无知的传播者有时竟成为大家公认的首创人。人们在不断传播和反复使用某一学说的过程中很可能会全部忘掉其不久以前的内容,这样就形成了历史上一再被删改、重写而找不出这种学说原始出处的情況。

就发生这种失掉来源出处的程度来说,它本身就是一个刚开始被研究的实验性问题。公开引证可能难以完全反映出科学著作的内在联系。随着知识影响的日渐加深,它会变得越来越难以察觉。这种影响很可能是通过接受某种理论观点及基本假设或标准的研究程序发生作用的。正如 Joshua Lederberg 曾在尤金·加菲尔德著的另一本书的前言中所指出的,“每个人都工作……就几乎根本没有人引证过”^[4]。专门的历史学与社会学研究可以用暗中引证补充公开引证。所谓暗中引证,就是用原文中的证据,如提到与某学说有关的人和带有某学说资料来源特征的术语等重新构成的引证。但是,我们还需要进一步观察,是否存在其他形式的认识影响,在无论是公开引证的还是暗中引证的参考文献和引文中都没有表现出来。

在引文分析的发展过程中,此类问题已开始受到它所要求的全面重视。例如 S. Cole、J. R. Cole、L. Dietrich、E. Garfield、M. V. Malin 及 H. Small 等人^[5]都对其进行了专门研究。H. Small 认为,即使失去了资料的出处,“重要性和影响都已得到承认的学说,很可能与一种或几种有时被大量引用的科学著作有关”。所以有人提出,引文计量不适用于平庸之才,因为此类资料出处的失去只出现在确实见到引文中有这种情况存在之后,这个条件是大多数已出版著作难以满足的。这一切提出了一个有趣的时间范围问题,因为在这个时间范围内引文分析能够有效地考证科学知识的源流。

读者将在本书中发现引文计量的很多其他用途,它们都已经被 SCI 利用。本书对其中一种最有争议的用途讨论得最多,即利用(我们中的一些人会将其说成滥用)引文计量作为评价研究工作和各个科学家发展潜力的主要的或决定性的依据。这是预料之中的。早在 1963 年,也就是 SCI 问世的那一年,就有人针对可能会有人“粗枝大叶地乱用定量引文数据去评价和挑选人员与学会会员”的情况提出过警告。而预先提出这一警告正是尤金·加菲尔德^[6]。他还进一步指出:“盲目地做出被引频次最多的著者就该得诺贝尔奖金的结论是荒谬可笑的。”

本书第十章根据过去 15 年的研究情况重新探讨了这个问题。阅读此章时,最好是把它作为那些敢于涉及上述危险领域者的方法性手册,而不要以为它是在为应用引文分析评价个人的科学工作进行新的辩护。该章重新提出这个问题是为了强烈地提醒我们,不能把引文计量作为评价个人工作的可靠的重要依据,因为被引频次充其量也只能帮助学识渊博的同行做出详细的判断。把引文计量用于比较不同的个人在科学贡献中的影响等一类活动,是一种有争议的做法。但是,这种做法为将这类比较活动置于有组织的怀疑之下提供一种极其特殊的机会,因为组织的怀疑是科学的基本特点之一。

凡阅读本书的读者都会发现书中想像丰富,情感洋溢。引文索引法成为科技目录学的一种标准方法已有 10 多年的历史了,但是它在社会学和历史学研究中的潜在作用大概还未完全被人们认识。它最突出的贡献之一是加速了文献计量学专业的诞生。文献计量学是在布拉德福(Bradford, S. C.)和 Gross 等学者的早期著作的基础上发展起来的,目的是要对科学的多重结构进行阐述和计量。现在从事这一领域研究的人越来越多,因此已开始轮到他们提出引文行为的基本问题了。

我可以这样说,本书对上述大部分问题都进行了认真的探讨。尤金·加菲尔德对科学的好奇心和怀疑态度导致他首先完成了对科学文献标引的概念性研究,接着又发明了一种科学文献标引的新方法。他把这种好奇心和怀疑态度用到发明 SCI 中,使他能与创建的 ISI 内的同事以及外单位的同事在研究引文分析基本原理的形成过程中进行广泛的合作。

参考文献

1. Kaplan, N. "The Norms of Citation Behavior: Prolegomena to the footnote." *American Documentation*, 16: 179 – 184, 1965.
2. Cole, J. R. and Cole, S. "Measuring the Quality of Sociological Research: Problems in the Use of the Science Citation Index." *American Sociologist*, 6: 23 – 29, 1973;
Moravcsik, M. J. and Murugesan, P. "Some Results on the Function and Quality of Citations." *Social Studies of Science*, 5: 86 – 92, 1975;
Chubin, D. E. and Moitra, S. D. "Content Analysis of References: Adjunct or Alternative to Citation Counting?" *Social Studies of Science*, 5: 423 – 441, 1975.
3. Merton, R. K. *Social Theory and Social Structure* (New York: The Free Press, 1968) Pp. 27 – 29, 35 – 38;
Garfield, E. "The Obliteration Phenomenon." *Current Contents*, No. 51/52: 5 – 7, (22 Dec. 1975);
Messerli, P. "Obliteration by Incorporation: Toward a Problematics, Theory and Metric of the Use of Scientific Literature." Columbia University, unpublished ms., 1978.
4. Lederberg, J. Foreword to Garfield, E. *Essays of an Information Scientist* (Philadelphia: ISI Press, 1977), p. xiii.
5. Cole, S., Cole, J. R. and Dietrich, L. "Measuring the Cognitive State of Scientific Disciplines." Pp. 221 – 229 and Garfield, E., Malin, M. V., and Small, H. "Citation Data as Science Indicators." Pp. 181 – 182 In Elkana, Y., Lederberg, J., Thackray, A., Merton, R. K., and Zuckerman, H. (eds.) *Toward a Metric of Science: The Advent of Science Indicators* (New York: John Wiley & Sons, 1978).
6. Garfield, E. "Citation Indexes in Sociological and Historical Research." *American Documentation* 14: 289 – 291, 1963.

前言

尤金·加菲尔德

从某种意义上讲,本书只是记述了用引文标引科学文献这种想法的发展过程。当 1963 年 SCI 年度本首次问世时,这种想法就变成了现实。与大多数明显打破传统思想或传统方法的新思想一样,SCI 的首次出现既合乎人们的情理和兴趣,同时也引起了激烈的争论。甚至在 10 多年后的今天,还有许多不爱利用参考文献的科学家对引文检索的逻辑感到陌生,或至少感到不大适应。在最初阶段虽然有相当多的人不爱利用引文索引和对它没有什么好感,但是在过去 10 多年中,接受和使用引文索引的人变得普遍起来。在我看来,任何一种新思想如此迅速地得到承认,都值得大书特书,写成为一本书。

我从 1955 年开始撰写有关科学文献引文索引的文章^[1],差不多用了整整 10 年时间,才使 SCI 的第一个年度本出版。当时,我已经写出了许多篇论文,足够将其汇编成好几百页的一本书。本书就是这些论文的副产品,其中不少论文都被我编著的《一个情报学家的论文集》一书摘录、重印或引用过^[2]。最初的计划是出版所有这些论文的重印本。但是,Robert M. Hayes (威利出版社的《情报学丛书》的编辑)认为,在同一主题的一系列单篇论文中,必然有许多不利于有效交流的冗余内容。他终于使我相信,如果有一本条理清楚、内容简洁的引文索引法方面的新书,情报科学的学生和实际工作者以及爱好文献工作技术的科学家的兴趣就会得到较好的满足。当我接受这种观点后,就决定用我的论文集作为本书的基本参考资料。

我对我的上述决定并不完全满意,因为这本书并不能完全像原始论文重印本那样起到真正历史记录的作用。虽然本书当中有一章讨论了引文索引法的历史,但是实际上它根本不是一部历史,而且从这本书中看不出我在大多数原始论文中喜欢使用的那种散文式文风。这是一个方面。另一方面,和原始论文集相比,本书的结构更严谨,引文更简练,对引文索引法及其应用的阐述更清晰。这些都是无可否认的。因此,我认为 Hayes 以前的预言是正确的,即对引文索引法感兴趣的大多数科学家来说,本书将会比论文集有用得多。对研究历史记录感兴趣的少数人来说,他们将在本书各章末尾的参考文献中找到该种记录的原始文献。

我提请那些特别想了解我是怎样产生科学文献引文索引想法的读者,注意阅读本书的第二章:“引文索引法的历史”。因为我在这一章中着力阐述了这种想法如何产生于几种特殊的知识爱好、专业问题和偶然事件的经过。就这一章确认的上述几种概念因素及其相互影响来说,它对科学引文索引之想法如何产生和上述概念又怎样导致 SCI 的研制,这些描述确实相当成功。但是重读该章之后我大吃一惊,因为我在该章中只字未提完成此项任务的一个非常关键的问题,即个人的入迷和专注。

从第二章的字里行间中,一部分读者可能会领悟出一种对各种挫折的难以明言的苦衷,如因传统技术的缺点,对这些缺点的普遍承认和在一般概念与经济上对各种新思想的抵制等所造成的挫折。我对这些挫折的反应是越来越坚定我个人的决心,首先研究了机器标引法,接着

研究了科学文献引文索引这一概念,最后终于使 SCI 变成了现实。这种专注使我 10 年来没有放松过对学术、专业和商务方面的追求,使我在不会有销售市场这种高明的预言面前敢冒创办一种新型索引的经济风险,使我敢于通过教育开辟 SCI 的市场,在美国科技信息研究所的投资上孤注一掷。

我提出专注这个问题不是为了标榜自己(事实上,我怀疑心理学家是否会称赞这种着迷的行为),而是为了发表个人的见解,因为这种见解可能有助于社会学家对科学发现过程的研究和有关理论的建立。有时也有一些人请我对下述一些问题发表意见,如建立 SCI 的“函数方式”是否势在必然;要是我没有研制出 SCI,其他人是否也能研制出这种索引等。此种问题一般当然是科学社会学中的典型问题之一。罗伯特·默顿(Robert K. Merton)对这个问题的剖析和回答要比我深入得多和透彻得多^[3]。事实上,他专门用 SCI 的研制,说明从一般概念模型中推导某一特定发现的困难^[4]。他发现这种模型存在着一些缺陷,只有弥补这些缺陷才能最后形成 SCI 的概念。另外,对于他的评论我还要补充一点:当一种新的发现出现时,只有通过某种思想入迷和专注的人,才能够完成从一般概念模型向特定概念模型发展,并使其从概念变为现实的任务。

还需在前言中指出,本书并没有对使用过引文分析的所有文献做出全面评价。这一方面是由于我决定主要以我自己的论文作为本书的基础;另一方面是由于用引文分析作为科学史和科学社会学研究工具的人在急剧增加。虽然我已经全面论述了引文分析正在被应用的所有领域,但是我相信并没有介绍所有的、正在使用引文分析的研究者的工作。这里谨向那些在我精心编织的大型参考文献网中被遗漏的著者表示我的歉意。如果此书将来再版,我将尽力使其更加全面。

参考文献

1. Garfield, E. "Citation Indexes for Science." *Science*, 122:108-111, 1955.
2. Garfield, E. *Essays of an Information Scientist* (Philadelphia: ISI Press, 1977). 2 Vols., 1314pp.
3. Merton, R. K. *The Sociology of Science*. Storer, N. W. (ed.) (Chicago: The University of Chicago Press, 1973). Pp. 343-370.
4. Merton, R. K. and Gaston, J., (eds.). *The Sociology of Science in Europe* (Carbondale, Illinois: Southern Illinois University Press and London: Feffer & Simons, Inc., 1977). Pp. 47-54.

目 次

中文版序言	尤金·加菲尔德(I)
英文版序言	罗伯特·默顿(III)
前言	尤金·加菲尔德(VII)
第一章 引文索引法的概念	(1)
第二章 引文索引法的历史	(5)
第三章 引文索引的设计与生产	(14)
第四章 引文索引法在专利文献中的应用	(27)
第五章 作为检索工具的引文索引	(30)
第六章 引文分析作为科学管理的工具	(53)
第七章 引文分析作为科学史的研究方法	(69)
第八章 引文分析用于科学结构的图示	(83)
第九章 科学期刊的引文分析	(124)
第十章 有关科学家的引文分析	(226)
结束语 引文索引法的未来	(236)
附录一 尤金·加菲尔德传略	(240)
附录二 新一代引文索引与学术信息资源整合体系 ——ISI Web of Knowledge	(243)
译后记	(247)

第一章 引文索引法的概念

引文索引的概念很简单。科技期刊的几乎所有论文、札记、述评、更正和通信,都包含引文。它们引用已发表的文献,一般给出题名、著者、出版地及出版时间,或者表示支持及提供背景,阐明或列举著者观点。引文是联系那些共同具有某些特定观点的论文之间的正式明显的纽带,引文索引就建立在这些纽带之上。它列举被引用的文献,指明引文的来源。任何查找文献的读者只要知道一篇被引用的文献,都可以发现一篇到十几篇关于这个主题的其他论文。而每一篇新查到的文献又提供了一系列用以继续查找的新的引文。

引文索引法的简易性是其主要长处之一。人们在编制一部传统的主题索引时,通常要具有关于该主题的专门知识,要阅读(或者至少要浏览)每一篇文献,并作出一系列的智力判断。这就涉及到选择描述文献内容的主题词。标引深度越大,即用来描述文献内容的词越多,标引员必须作出的智力判断就越多。这些判断费时费工,会很快花去标引工作的经费,或许还会影响完成索引的及时性。主题索引滞后于文献出版一年或一年多才问世,是很平常的。这种状况限制了主题索引作为文献检索工具的效果。

花在标引深度上的费用甚至更为要紧。虽然大多数重要的主题索引都是由联邦政府财政支持或资助的,因而没有把全部生产费用转加在它们的订户身上,但是这些主题索引仍必须权衡标引深度和标引成本。

近年来一些机构采用一种叫作“篇名词标引”的方法,作为权衡标引深度和标引成本的一种方式。采用这种主题标引方式,文献标引用词是文献著者在篇名中所使用的词。标引员无需做出智力判断,出现在篇名中所有的词都被认为是可以接受的标引词。因而,就可以比传统的主题索引更快、有时更省地编制篇名词索引。

然而在篇名词索引编制的速度和经济性方面也存在着一些问题。其一是篇名的有限长度限制了标引时可用词的数量。这样编出的索引标引深度很浅,往往只提示出论文的中心主题,而忽略了论文所包含的、附属于这些中心主题的所有资料,这些资料对索引的用户或许是重要的。其二是著者们在拟定准确地描述文献内容的篇名方面的能力是极不平衡的,甚至带有倾向性。因而,篇名词标引用词的质量也是极不平衡的。

现在有一些弥补这些不足之处的方法。在一些篇名词索引中,通过增补一些由标引员赋予的主题词来提高标引深度。有些篇名词索引则通过在上下文中显示篇名的每一个词来改进索引的质量。

引文索引用著者的引文来代替标引员的判断,从而解决了标引深度与标引成本的权衡问题。这种方法的优点是无需智力标引,不必在标引深度与标引词的质量二者的任何一方作出妥协。因为在一般情况下,参考文献的长度是不加限制的,所以对引文索引的标引深度也没有人为的限制。虽然不同的著者提供的参考文献的质量不同,不同类型的论文提供的参考文献的质量也不相同(例如,述评文章的参考文献一般比其他类型的文章的文献更为全面),但是,优秀的科学文献的评定标准以及把良好的参考文献视作学术水平的标志的做法,使得引文质

量(就标引的目的来看)比篇名的质量要高得多,也齐整得多。

引文索引的另一重要长处是它的检索效果。引文索引法的这种特性有两个组成部分。一是查获量,关系到查获尽可能多的相关文献;二是检索效率,关系到把文献检索者在识别相关文献时所必须筛选的不相关文献降低到最低限度。

标引深度是检索效果的主要数量测度。标引用词用得越多,文献主题就被描述得越细。随着标引深度提高,检索人员满足其检索需要的可能性也会增加。因为每篇科技论文一般含有大约 15 篇引文,所以引文索引的平均标引深度为 15。传统的主题索引大多不能达到这样的深度。

检索效果也有其质量方面的问题,它涉及到一个标引用词如何准确、全面地描述相关文献。对文献描述的准确性是语义问题,在主题索引中,它会产生一系列的问题。基本的问题是不同人的用词习惯也不相同。不管标引员多么有才能,但是显然不可能妥善地协调人们在用词方面的差异,以选出一系列主题词来,使它们忠实可靠地把科技文献中的复杂概念传递给正在查找它的人。

语言的动态性使这一工作变得更为困难。新词被不断地采用,旧词被不断地淘汰,有些旧词被赋予新的含义。

另外一个问题是主题索引的用词需要标准化,以便对标引的一致性和质量实行某种程度的控制。这样,标引员在处理前两个问题时就会减弱语言的丰富性及多样性。

上述所有做法的结果是,使传统主题索引的检索过程更复杂了,查出的文献更少了,检索效率更低了。

把引文用作标引用语,由于避免了语义方面的问题,也就提供了一些主题索引所丧失的手段,使索引查找简便易行,查获量大,效率高。一条简单的引文“Fisher, M. E. 《数学和物理学杂志》,1964 年 5 期第 944 页(Math. Phys., 5, 944, 1964)”,会把检索者直接指向引用过这篇重要论文的一个论文目录。经验表明,引用的论文中有很高比例是相关的。检索者无需判定标引员最有可能使用哪个主题词来描述相关的论文。检索者的语言习惯也不会影响检索结果,自 Fisher 的那篇论文发表后科技术语发生的任何变化也不会影响检索结果。

换句话说,引文是一个主题的准确、不含混的表述,无需解释,不受术语变化的影响。另外,引文在任何时候都将保持它的准确性。它也可以用于不同语种的文献之中。一系列实例已经最好地证明了这种语义的稳定性和准确性对检索过程的重要性。

准确性直接影响检索的效率。例如,如果某人用 1971 年的《医学索引》(Index Medicus)的主题索引查找有关阿斯匹林对前列腺素产生影响的资料,“前列腺素”是理想的检索出发点。这个词把检索者指引到一系列不同的副标题之下的几百篇论文。最切合查找主题的副标题是“拮抗剂、抑制剂”和“生物合成”。在前两个副标题下分别列有 10 篇及 9 篇论文。在这两个副标题下都分别收录了两篇似乎是论述阿斯匹林与前列腺素关系的论文。但是,检索者要检出这四篇相关的论文,必须阅读这 19 篇论文的篇名。与此相反,查找 1971 年的引文索引,在“Vane, J. R. Nature New Biology, 231-231, 1971”之下,一下子就可以找到一个由 15 篇论文组成的目录。因为这些论文全都引用了一篇有关“阿斯匹林类药物抑制前列腺合成的作用机制”的论文,这正是所要查找的确切主题,这个由 15 篇论文组成的目录包含相关文献的比例会是很高的。

语义准确性也能够影响检索的查获量。1963 年发表了一篇有关“出生的季节性变化

(Seasonal Variation in Birth)”的重要论文。任何想要查找有关该主题资料的读者在 1965 年的《医学索引》中,是不可能找到这篇重要论文的,除非这个读者是在一个不大可能问津的标题“周期性(Periodicity)”之下进行查找。

查找“人种改良(Euphenics)”这个主题,可以说明语义的稳定性对查获量的重要性。J. Lederberg 教授 1963 年在《自然》杂志上发表的一篇论文中首先采用了这个词。当时人们一直使用“人类发展工程(Engineering Human Development)”这个词来表示这个主题,至今仍有人使用这个词。如果对主题索引中的“人种改良”进行查找,尽管它仍是一个新词,仍可以查到一些使用这个新词的论文,但是使用“人类发展工程”这个词的论文就会遗漏。与此相反,在主题索引“人类发展工程”这一主题之下进行查找,就可能会遗漏使用“人种改良”这个词的论文。然而,在引文索引中,凡是引用了 Lederberg 教授这篇重要论文的各篇论文,都将会集中在一起,不管著者采用什么术语来描述这个主题,即使人们用两个词来查找主题索引,但是它的检索效率和引文索引的检索效率之间有着明显的差别。对主题索引进行完整的检索,需要查找两次;完整地检索引文索引,仅仅需要一次。当表达该主题的术语继续发展为“遗传工程(Genetic Engineering)”、“质体工程(Plasmid Engineering)”和“重组脱氧核酸(Recombinant DNA)”时,该主题领域重要论文语义的稳定性^[1]将使得引文索引检索的效果更加明显。

一个标引词描述相关文献的综合程度,反映了学科和时代的局限性。主题索引收录了一个特定时间范围内的一个特定学科或一组学科的文献,如化学、医学或生物学。这使得标引词的范围受到了学科和时代的局限。引文索引则不然。

即使为一个特定时间范围内发表的若干学科的文獻编制引文索引,作为标引用词的引文也丝毫不受这些限制。著者们总是一致地引用他们所从事的学科以外的论文,引文的范围跨越整个 20 世纪科学的广大领域。把这些引文用作标引用词,能够使引文索引反映科学发展的曲折的轨迹,因为引文交叉往来于不同学科之间,跳跃往返于不同时代。这种特性大大地提高了引文索引检索文献的能力。

一个实例相当惊人地说明了引文索引中学科交叉的程度。从 1961 年至 1965 年,SCI 把引用了爱因斯坦在 1906 年《物理学年报》上的一篇经典论文的所有文章都联系起来,发现这些文章来自下列期刊:

乳品科学杂志(Journal of Dairy Science)
美国化学会会志(Journal of the Chemical Society)
聚合物科学杂志(Journal of Polymer Science)
药学与药理学杂志(Journal of Pharmacy and Pharmacology)
比较生化和生理学(Comparative Biochemistry and Physiology)
普通生理学杂志(Journal of General Physiology)
国际工程科学杂志(International Journal of Engineering Science)
材料杂志(Journal of Material)
水污染控制联合会会刊(Journal of the Water Pollution Control Federation)
美国陶瓷学通讯(American Ceramic Society Bulletin)
美国声学会会刊(Journal of the Acoustical Society of American)
化工科学(Chemical Engineering Science)
工业与工程化学处理的设计与发展(Industrial Engineering Chemistry Process Design and

Development)

胶体与界面科学杂志(Journal of Colloid and Interface Science)

流体力学杂志(Journal of Fluid Mechanics)

润滑技术杂志(Journal of Lubrication Technology)

分子生物学杂志(Journal Molecular Biology)

食品科学杂志(Journal of Food Science)

生物化学杂志(Journal of Biological Chemistry)

沉积岩石学杂志(Journal of Sedimentary Petrology)

科学仪器评论(Review of Scientific Instruments)

电化学学会会刊(Journal of the Electrochemical Society)

引文索引法的时间外伸幅度与检索的查获量同等重要。以上述爱因斯坦 1906 年的论文为例,引文索引能够从任何时间查起,只要那时发表该主题的相关论文。如果那篇论文在索引所覆盖的年份之中,那么索引将会把最新发表的文献指引给检索者。然后又可以用最新的文献循环检索,查出它们的引文的发表时间,这些引文又将给检索者指出一批近期的文献。检索者采用这种方式,往往只使用一期引文索引就可以获取有关一个主题的历史发展的文献。对于主题索引来说,这就需要查阅许多期,而且还要方法得当。

引文索引在时间上从过去的文献指向新近发表文献的跳跃往返的能力,查找交叉学科发展的能力,克服由于语义问题所引起的检索局限性和复杂性的能力,在切实可行的时间和经费的情况下给文献提供深度标引的能力,都已经证明它是既有实践意义,同时又具有理论价值的。尽管引文索引标新立异摆脱了传统的标引方法,但是在十年之中,引文索引作为一种进行科技文献回溯检索的重要方法已经得到了承认。

引文索引法的影响甚至已经超出了这些。它在近期文献通报服务、图书馆管理以及科学政策、科学史、科学文献等方面的研究中已经发挥了重要的作用。

参考文献

1. Garfield, E. "Genetic Engineering - Too Dangerous to Continue or Too Important to Discontinue?" In *Essays of an Information Scientist*, Vol. 2 (Philadelphia, ISI Press, 1977). Pp. 335 - 341.

第二章 引文索引法的历史

与其他许多新事物一样,用引文标引科学文献的概念也是在为实现其他目的的活动中发展起来的。第二次世界大战刚一结束,联邦政府的研究与开发经费就开始出现急剧上升的趋势,而且这种趋势持续 20 年之久。政府的这种经费刺激产生了两种结果,一是加快了科学研究的步伐;二是报道科学研究活动的科学文献急剧增加。由于政府考虑到科学家之间的情报交流系统无力处理日渐增多的文献,所以资助了许多项目,旨在发展和提高科学情报传播与管理的手段和方法。例如,武装部队医学图书馆(即美国国家医学图书馆的前身)资助约翰斯·霍普金斯·韦尔奇医学图书馆进行一个研究项目,以研究有关机器在生产和编制医学文献索引中可能起到的作用^[1]。作为该研究项目小组的一个成员,我逐渐对是否能用机器和怎样用机器提取有效地描述文献内容的标引词产生了浓厚的兴趣,这种标引词的产生应当无需标引人员进行智力判断,旨在取消人工标引的引文索引的概念便是这一兴趣的产物。研究的结果证明,编制索引本来不需要机器。

零星的经验

科学文献引文索引的概念综合了通过几年时间积累起来的零星经验和看法,这种积累是从我参加韦尔奇医学图书馆项目的两年里开始的。通过对各种索引的调查,我进一步了解了标引过程中智力判断的性质以及索引与其用户之间的功能关系,还了解了字顺主题索引与分类主题索引的结构与功能特点。在我自愿为《化学文摘》杂志编写文摘期间,我发现文摘利用率的增加是因为其参考文献中收录了被著者引用过的主要论文的文摘。我发现这个问题后,就建议在《化学文摘》中参见栏目的基础上发展出一个参考文献索引。更重要的是,在 C. D. Leake 的建议下,我对综述进行了一次专门研究。综述在为科学家提供文献指南方面起着特殊的作用。我很快发现,这种综述性文章几乎为每个句子提供一条参考文献,而每个句子本身又是被引用文献的非常明确的标引用词。1953 年我从《谢泼德引文》(Shepard's Citations)的前任副社长 William C. Adair 那里得知,有一种使用引文的法律诉讼文献索引。

在现有的几种主要的引文索引中,《谢泼德引文》的历史最长,从 1873 年起成为法律专业检索法律判决书的一种工具。它起到这种作用是由于它把联邦法院和州级法院以及各种联邦行政机构所判案件中使用过的判例引文编成了一览表。这种法律的“引用者”系统为如何使引文索引成为一种有效检索工具提供了模式。但是要指出的是,当时的《谢泼德引文》并不是今天这样的法律期刊文献索引。

其他零碎经验则是来自于我后几年对图书馆学和结构语言学的正式研究。开始时,我觉得机器标引中要解决的基本问题,似乎是识别论文中的关键词,并使其串连在一起构成有用的标引词或标引短语。但是,对各种不同文献,尤其是化学与医学专业文献及其索引的研究表明,问题要比上述看法复杂得多。