

科学引文分析

许志强

(西昌学院 图书馆, 四川 西昌 615022)

【摘要】文章概述引文、引文索引与引文分析 科学引文分析与科学引文分析的应用。

【关键词】引文; 引文索引; 引文分析; 科学引文分析; 应用

【中图分类号】C7 【文献标识码】A 【文章编号】1008-6307(2004)02-0127-04

The Analysis of Science Quotation

XU Zhi-qiang

(Xichang College, Xichang, Sichuan 615022)

Abstract: This paper defines the quotation, quotation index, quotation analysis and the analysis of science quotation. It further illustrates the application of the analysis of Science quotation.

key words: quotation; quotation Index; quotation analysis; the analysis of science quotation; apply

一 引文、引文索引与引文分析

1. 引文

“引文”来源于英文“Citation”, 国外习惯于称“引文”为再版目录, 是一种与一次文献同时出版的“准二次文献”。具体地讲“引文”就是从一部著作引用某一章节, 对于该一著作的或对于一项陈述或主张作为权威依据的某些来源资料的参考注释”。这是一种简短的参考性的记载, 用以指明引用的文字, 公式, 结论等的来源, 或指明含有类似资料的出版物章、节。

由于科学劳动的继续性和科技论文的整体性, 所以引文是普遍存在的。马克思讲: “科学研究部分地以今人的协作为条件, 部分地又是以对前人劳动的利用为条件”。这里“前人劳动者的成果”指的就是历史上留下的图书文献资料。因此“科学研究必须充分地占有资料”, 必须“站在巨人的肩膀上”。正如英国著名物理学家齐曼教授在《知识的力量》一书中所指出的: “一篇典型的科学论文中, 往往大量地参考或引用其他人的实验、计算、观测或理论。它不是凭自己想象向未知进军, 而是根据前人有所把握的研究, 小心翼翼地, 一步步地前进”。据统计, 全世界

约有90%以上的科学论文附设有引文, 平均每篇文献具有引文15篇。我国目前88%左右的重要科学论文带有引文, 平均每篇中文科学论文有引文8、9篇, 对部分中科院学部委员的引文统计, 他们的论文平均引文量约为12篇。目前西文图书附书末索引的达94%, 日文图书达到90%, 中文图书达到20%左右。

随着引文(参考文献)著录标准化、规范化, 引文(参考文献)已成为科技文献的重要特征, 引用已成为文献之间的一种普遍现象。根据引文在正文中的位置, 可分为文内引文、页下引文和末尾引文(将引文放在篇、章或书的末尾, 即参考文献)。研究表明, 一篇论文的质量越高, 越容易引起学术界的注意, 因而在相关论文中被引用的概率就越大, 次数越多, 社会响应值越大。反之亦然。

2. 引文索引

引文索引是建立在文献的引用和被引用关系上的一种新型索引, 是从被引用文献去检索引用过该文献的文献索引。严格地讲“引文索引”是指“已发表文献的一种索引, 在书目中所引用的每一篇早期文献附在正被标引的近期文献中”。美国著名情报学家加菲尔德在力著《引文索引》第一章“引文索引法”的概念评述”中, 把引文索引和主题索引、篇名索引

收稿日期: 2004-05-28

作者简介: 许志强(1962—), 男, 研究馆员, 主要从事信息(情报)管理研究。

等方法做了比较,指出它最大的两个优点——编制简易、检索效果大。他说:“引文索引就是建立在这些纽带基础上的。引文索引法的简易性是它的主要长处之一。引文索引的另一个重要优点是它的检索效果。……引文索引法的影响甚至已超出了这些,它已在现期文献通报服务、图书馆管理以及科学政策、科学史、科学方法等的研究中起了重要作用。”美国著名学者温伯格在其报告中指出:“鉴于标引是建立在作者评价的基础上,而不是建立在一位编辑者评价的基础上,那么就某篇论文与其它论文的关系来讲,引文索引更适合于用户的观点,而不是标引员的观点。”他讲:“我们想提请技术人员注意这种新的检索工具,尽管较少使用,但这种工具是非常有效的。”匈牙利科学家布劳温在“科学计量学与文献计量学”一文中指出:“尽管《科学引文索引》在开始构思创立的时候,只是为了文献检索之用,但不久人们即发现,该检索系统所拥有的应用于科学计量学研究的能力是相当广泛的。”

3. 引文分析

引文分析(Citation Analysis)就是利用图论、弗晰数学、数理统计及其它数学、逻辑思维方法,对科技文献的引用或被引用现象和规律进行分析,以便揭示出它们所蕴含着的研究对象具有的特征或者对象之间的关系。从表面上看,著文与引文的关系是相当简单的引用与被引用的关系。然而引文的内涵却是丰富多彩的,著者的引用心理、情报行为却是千差万别的。实际上,科学论文的引用就是对已有科学成果的回采,著文与引文的表现形式是时间序列、聚类。加菲尔德在《引文索引它的理论和在科学、技术、人文中的应用》(New York, John Wiley & Sons, 1979, P81-147)中分析文献被引用原因,不下十五种:

寄先驱者学术成就以敬意,对已有成果表示赞赏与褒扬;核对已经应用过的方法和仪器;提供著文以外的,更为广泛的资料;对自己以前工作的修正、补充和深化;对他人的工作加以反驳、批判与更正;评论前人工作的成效、优点与缺点、成就与不足;为自己的主张提供文献依据和支持;展望前景;对不良的传播、不适当的引用和该引而未引的文献进行澄清;鉴定实验数据和理化常数;核对引文中某一个主张或概念、业已被讨论过;核对原始资料中或其它著作中的起因人物的一个概念或一个名词;承认他人的工作或观点;对别人要求优先权提出争议。

对上述各种各样的引用心理,进行由表及里的

分析,揭示引文形式与内容之间的辩证关系,揭示著者引用文献的心理规律(求快心理、求实心理、求新心理、求近心理、选择心理、习惯心理、创造心理),发展引文分析方法。由于引文的外表特征,使得我们可以实施由近及远、由表及里的内容分析。这种独特的分析方法,不仅容易抓住实质,而且工作效率较高。

引文分析是文献计量学、情报计量学、科学计量学的重要组成部分,是二门学科的重要研究课题,离开了定量研究,引文分析就失去了意义。因此,引文分析便于量化是引文分析的重要特性,是引文分析方法的重要手段。加菲尔德对1981—1982年生命科学被引次数最多的论文104篇进行了统计分析,其结果为:109篇论文两年内平均每篇被引用68次,1981年平均11次,1982年平均57次,被引最高者为156次,最低者为48次。

随着科学技术的迅猛发展,科学的结构越来越复杂。为了揭示科学结构之规律,世界科学家进行了坚持不懈地努力,作出了杰出的贡献。科学学创使人贝尔纳曾说过:“科学上的一般发展模式是相当明确的,与其说它像一棵树,不如说它更像网。直接与问题或应用有关的科学工作的内容可以比作网孔,网线的交叉意味着经验和思想的汇集,并且是产生各种技术和学科的焦点”。加菲尔德·E曾绘制过引用时序网络图。实际上,在加菲尔德编制引用时序网络图之前,著名科学史家普赖斯就曾做过类似尝试,编制过“基因代码”时序关联图。这种网络图具有编制简单易行,形象直观、便于计算机处理等特点,因此深受人们的欢迎。这正是英国著名情报学者布鲁克斯·B·C所追求的“知识地图”。除了引用时序网状图外,还可绘制静态引用网状图、动态引用等高线图、框图、二维空间图等。

二 科学引文分析

最早的引文分析可追溯到1927年,格罗斯对《化学教育》期刊上刊登文章的引用文献进行了研究,把引用文献所在期刊通过其上的文章被引用频次排序,便得到一个期刊顺序表。由于期刊登载文章被引用频次的不同,各种被引用的期刊,对《化学教育》来说,重要性也就不一样。随后,有关引文分析的文献越来越多,莱恩和山笛松提出的文献老化问题,布克兰特(Buckland, M.K)的老化分散问题,利普茨(Lipetz),贝一埃米(Ben-Ami)的关于选择性改革问

题都是基于引文统计的数据进行分析的。真正对引文分析有承上启下贡献的是美国费城科学情报所所长,著名情报学家加菲尔德。第二次世界大战期间,加菲尔德·E当了几年滑雪兵,战后,享受政府对复员军人的优待,进入哥伦比亚大学学习化学。同时兼任出租汽车司机。在这期间,他还为化学文摘社(CAS)作过义务文摘员。在工作中,他认识到检索系统的优越性,开始对检索系统产生了浓厚兴趣,一个偶然机会,他听说,约翰·霍普金斯(Johns Hopkins)大学正要开始一项医学文献的索引计划。在1953年召开的一次科学文献机器管理讨论会上,加菲尔德透露了这个想法。这个想法引起了已退休的某公司副经理阿德尔(Adair·W·C)的注意,他写信建议加菲尔德参考本公司从1873年开始出版的一种法律业务工具书“谢泼德引文”(Shepard's citation)。

“谢泼德引文”以联邦法院和州立法院的案件为条目,在每个案件下,列出了审理此案所参照的法律条文,所参考的其它案例,以及其它对律师有用的材料。用引用和被引用的关系,按案例和法律条文联成了一张网络。如下所示:

被引用的案例→101 Mass 210

a 130 Mass 65

q 192 Mass 69

o 221 Mass 310

他讲:“在我的事业中,这是一个决定性的时刻——从这个时刻开始,我确实知道了我该做些什么”。1954年,加菲尔德在哥伦比亚大学获得了图书馆学硕士学位,第二年,他的论文“引文索引用于科学。在美国《科学》(NO:122,P108-111)杂志上发表。论文第一次提出了用引文索引检索科学文献的方式(滚雪球式检索),向独霸情报检索的主题分类法提出了挑战。1958年,加菲尔德租来了一台制版机和一台小型胶印机,在一间经过改建的鸡棚里开始印刷他的第一种情报出版物活页的《现刊题录》(生命科学)。这样,费城斯普林加尔登大街上诞生了加菲尔德协会。在这期间,加菲尔德协会为制药公司搞过一种专利引文索引系统。后来又出版了“化学文献索引”。1960年4月,加菲尔德登出了广告,宣告“科学情报研究所”成立。1961年,研究所得国家健康研究所的支持,后改由国家科学基金会(National Science Foundation)资助,用计算机列举出89万篇著文的130余万条引文,编制成《1916年遗传学引文索引》,为遗传学文献发展一种试验性的引文索引系统,并探索

多学科引文索引的可行性。同年,加菲尔德获得了宾夕法尼亚大学结构语言学哲学博士(ph.D),1963年,包括各种学科的“科学引文索引”(SCI)单卷本终于问世了,1965年,加菲尔德在IBM-707计算机上运用引文索引方法,曾经构成主要科学思想网络联系,这些思想导致20世纪遗传学上最大的事件——在脱氧核糖核酸分子中译解出遗传密码。同年,著名科学家普赖斯(Price,D.J.D)在《科学》上发表了“科学论文的网络”(Networks of Scientific papers)Vol149,p510-515。1970年,加菲尔德率先提出了引文分析测度“影响因子”(impact factor)1973年,研究所开始出版“社会科学引文索引”(SSCI),1978年又出版了“艺术和人文科学引文索引”(A&HCI)。为了深化引文分析方法,在1978年匈牙利出版了一份国际权威刊物《科学计量学》,主编为布劳温博士,加菲尔德博士,由荷兰阿姆斯特丹的埃尔塞维尔科学出版公司和匈牙利科学院出版社共同出版。该刊有一个庞大的国际编委会,我国学者赵红州是该刊国际编委。这一年10月17日,耗资650万美元的科学情报研究所(ISI)新国际总部的开工典礼在费城大学城科学中心隆重举行。这一期间,加菲尔德取得了许多研究成果,对科学引文分析作出了重大贡献。1983年加菲尔德曾访问过我国,并在我国设立了办事处,办事处主任为陈华江博士。1984年12月20日,奥列根大学科学研究所的摩拉夫西克把曾赖斯纪念奖的奖状和奖章授予加菲尔德,使他成为该纪念奖的第一位得主。

在加菲尔德引文分析的基础上,美国著名科学家、科学史家和科学计量的奠基人曾赖斯(Price,D)发展了引文分析方法,提出了科学论文的网络模型。曾赖斯一生著述甚丰,一生出版专著14种,发表论文240篇。加菲尔德使用自己主编的自然科学、社会科学、人文科学引文索引(SCI,SSCI,A&HCI),对曾赖斯的全部著述实施引文分析,结果表明:截至1983年底,至少有2200件以上的文献引用了曾赖斯论著。加菲尔德在题为“献给曾赖斯”的演讲中指出:“我们的生命是有限的,如果在活着的时候继续研究这一领域的话,就不能不每天都要想起曾赖斯给予的影响。因此,我们不必悲痛他的消逝,曾赖斯是不死的”。

为了全面、准确地反映美国的科学生产率,提高科研管理水平,美国国家基金会(NSF)还利用引文分析编辑了《科学指标》,并且规定每两年向国会报告一次。

随后,美国学者怀特·H(White,Howard)和格里菲什·B(Griffith,Belrer)根据对情报科学主要作者工作的引文分析绘制了“情报科学主要作者图”,在此基础上,英国学者维克利·B.C等完善了此图。

80年代以来,引文分析获得了重大的发展,加菲尔德创立的科学情报研究所(ISI)也有了比较大的发展。作为科学机构,它拥有100多名科学家(包括两名诺贝尔奖获得者),作为商业机构,它为用户提供20多种出版物,雇员近700人,在9个国家设立了办事处,年销售额高达3000多万美元。在多伦多、伦敦、巴黎、波恩、堪培拉、新加坡、东京、香港、台湾地区等地设立了分公司及代理商。

三 科学引文分析的应用

1. 预测科学技术的发展趋势

科研是科学技术发展的主要动力,学科发展新动向、新趋势不仅是未来学(预测学)研究的重要课题,也是科学引文分析的应用之一。

科学哲学大师库恩在《科学革命的结构》一书中指出:“科学文献所援引的专门文献的变化,可以看做是科学革命可能的征兆”。为此,加菲尔德从科学论文的引用与被引用现象入手,利用科学论文的被引用频次来预测科学的发展趋势。他从1982年40多个国家和地区的3000多种刊物,100多万著者的540000篇文献中选出了被引用频次最高的50篇文献,这50篇文献发表于32种期刊上。通过分析知道,32种期刊中涉及生命科学的期刊有17种,共计文献30篇。其中《生物化学杂志》载文量最多,共计9篇,被引文献的生命科学领域论文遥遥领先。由此可见,生物技术是六大群体技术之一,是信息时代的带头学科。这与实际情况是完全一致的。然而值得注意的是:据科学引文索引统计,我国生物学论文却由1983年的343篇下降到了1988年的287篇,在我国的论文中所占比例由16%下降到4.7%,其下降幅度之大,应引起我们的高度重视。

2. 衡量个体、团体乃至国家的情报能力

所谓情报能力就是情报生产、情报交流和情报利用能力。如果从科学引文分析的角度来看,情报能力不外乎包括:利用能力和被利用能力两大类。基于此,加菲尔德将1973、78年《科学引文索引》(SCI)的数据加以整理,并对此进行了统计分析,其结果为:瑞典的被引用百分比为75%,丹麦的被引用百分比

为73%,荷兰的被引用百分比为72%,以色列的被引用百分比为71%,加拿大、挪威、芬兰的被引用百分比均为70%……。中国科技情报所利用《科学引文索引》(SCI)、《科学评论索引》(ISR)、《科技会议录索引》(ISTP)、《工程索引》(EI)对我国科技人员发表论文情况进行了调查,1988年,SCI、ISTP、ISR、EI四种系统共收录科技论文949,074篇,其中我国科技论文11,854篇,占世界论文总数的1.25%,居世界第14位。据此,我们就可以判断上述国家情报能力的高低,并能对上述国家的情报能力排队。确定“核心期刊”

期刊是交流科技情报的主要文献,确定高质量的权威期刊——“核心期刊”,对于节约收藏空间、节省管理费用,以及帮助用户精选阅读期刊,节省用户的精力和时间,具有重大的意义。“核心期刊”的确定方法很多,其中引文分析法就是一法。1972年,加菲尔德在《科学》(Science)杂志上发表了引文分析方面的研究文章,他用1969年一个季度内各种期刊被引文献的统计得到核心期刊152种,这152种期刊包含了当时《科学引文索引》所用2000种期刊总数被引文数的50%。1976年,加菲尔德又在《自然》(Nature)杂志上发表了1974年统计出来的核心期刊206种。1979年,加菲尔德把《科学引文索引》的3722种期刊,按其被引频次高低,排成了一个顺序表。有的学者认为:综合性图书馆按上述被引次数多少而排的顺序表中,取前500—1000种期刊为核心期刊就足够了。日本学者绿川信之对上表加以改造,重新构造了数学等六个专业的期刊表。

3. 研究著者的分布规律

科学已由“小科学”时代跨入了“大科学”时代,当代科学劳动是有组织的协同作战,科学引文分析可以研究科学劳动的结构,揭示科学家的成长规律。加菲尔德曾利用《科学引文索引》(SCI)选出1981、1982年被引频次最高的104篇文章,统计这些文章的著者,其中署名3人的论文最多,量高的为14人。中国科技情报研究所的统计表明:1988年,SCI、ISTP、ISR、EI四系统共收录我国61,102个作者发表的12,392篇论文,以第一作者计,平均每人1,12篇。其中5篇以上论文的作者16人,5篇论文的作者11人。

4. 科学引文分析用于人才评价

首先,我们假设:“如果文献被引用的次数越多,其质量越高,其作者的水平也高”。由于引文的价值性与差异性,使引文分析用于人才评价(下转137页)

片、广告节目、片头等的制作。

利用非线性编辑系统进行后期编辑是一项综合性很强的工作,不但要求编辑制作人员要有很扎实的画面编辑能力、视听语言能力和一定的音乐素质等,还要求制作人员对编辑时用到的软硬件系统、不同格式的多媒体素材等都要十分熟悉,这样才能在节目制作中做到得心应手;才能把熟练地前期拍摄的素材进行剪辑组合,配上解说词、音乐、效果,使声

画合为一体,最后编辑成为主题思想完整、画面镜头语言流畅的电视节目(电视片)。

由于非线性编辑系统依赖于计算机,所以在运用它进行视频编辑时会遇到不少的问题,需要我们在工作中不断地总结经验教训,勇于创新,勤于思索,在创作中逐步提高,为制作出优秀的电视节目而努力奋斗。

注释及参考文献:

- [1]黄亚安.电视编辑[M].复旦大学出版社,1991年版
- [2]刘旭,杜筱欧.关于特技在电视专题片中应用[M].西部广播电视学刊,2003.3
- [3]张冠文,王松丽.论影视作品中动画形象的创作[J].电视字幕.特技与动画,2003.2
- [4]张哲西.电视专题节目制作知识[M].四川人民出版社,1992年版

(上接130页)成为可能。加菲尔德博士和 I.H.Sher 曾经用该索引 1961 年版本,来研究 1962 年及 1963 年两年中,在物理学、化学及医学领域内的诺贝尔奖金获得者在授奖前发表的科学论文是否比一般科学工作者的被引率高些,结果表明:获奖者的论文每人平均被引次数是 169 次,而一般科学工作者平均被引次数 5.51 次,两者的差距是相当大的。为了排除获奖者论文发表多些的可能。加菲尔德用单篇论文平均统计,结果为:获奖者平均每篇论文被引次数为 2.9 次,而一般科学工作者则是 1.57 次。

注释及参考文献:

- [1]加菲尔德·E引文分析作为一种工具用于期刊评价[J].科学Vol,178.No.4060、1972
- [2]加菲尔德·E,sher J.H.利用引文索引来评价科学文献的新因素[J].美国文献工作Vol,14,1963
- [3]加菲尔德·E,引文索引用于科学[J].科学No.122.1955
- [4]加菲尔德·E,引文索引[J].纽约.约翰.威利公司.1979
- [5]加菲尔德·E.Is citation analysis a legitimate evaluation tool? Scientometrics 1979
- [6]加菲尔德.献给普赖斯[Z].1984.12.20.在获普赖斯纪念奖的演讲稿
- [7]缪其浩.加菲尔德和引文索引[J].情报科学.1881.1
- [8]普赖斯.科学论文的网络[J].科学.No.149、1965

1977年,扩大了统计范围,对1950、1977年在物理、化学和医学领域内162名诺贝尔奖获得者的论文被引数作了全面统计,计算年为1961—1975年共十五年。

结果表明:荣誉称号高的学者的文献被引次数较多,诺贝尔奖获得者的平均年龄为37岁,这与科学家的最佳发明年龄35—45岁是一致的。用引文分析来评价人才在国际上已引起重视,世界各国的科学家正进一步完善引文分析理论,并运用于实践中。