

叶新明

引文中伪引和漏引的机制分析

ABSTRACT After discussing false citations and missing citations as well as their rates and reasons, the author proposes some methods to prevent them. 1 fig 1 tab 10 refs

KEY WORDS Citation analysis Relevant documents

CLASS NUMBER G257.5

科学文献的引文早在18世纪初就出现了。但在我国近几年才越来越重视。它不仅已成为学术论文的组成部分,同时也是衡量学术论著、学术期刊和论文作者学术水平的重要指标。正因为如此,也往往产生一些伪引和漏引现象。笔者调查分析了伪引和漏引的现象及其产生的原因,并提出一些改进建议,以求教于学界同仁。

1 引文中的伪引和漏引

引文是为撰写或编辑论著而引用和参考的有关文献,通常列于文后,又称参考文献。它的显著特征是内容(或主题)的相关性,就是指能够用来支持作者论点的那些相关文献。但是,作者在实际撰写或编辑论著时,由于动机的复杂性,致使在文后列出一些与本文不相关的文献或者漏列了一些已经引用的相关文献,从而造成了事实上的伪引和漏引现象。

1.1 伪引和伪引率

对论著的作者在引文中列出的非相关文献,我们称之为伪引。为了准确地反映这种现象的存在程度,本文导入伪引率的概念。所谓伪引率(P_r),就是指引文中非相关的文献数量占引文数量的百分比。凯诺曾对结构工程领域的21位杰出科学家最近发表的42篇(每人2篇)论文所参考的344篇文献,按实质引用、有力引用、适度引用和表面引用进行过分析^[1]。后来有人对其结果进行模糊综合评价后得出:22.3%的引文为实质引用和有力引用,27.7%的引文为适度引用和表面引用^[2],这就是说,无关的引用占50%。这一结果足以说明伪引存在的严重程度。

1.2 漏引和漏引率

论著的作者有时在引文中只列出曾经引用的部

分相关文献,而另一部分曾经引用而未列出的相关文献,我们称之为漏引文献。这里强调“曾经引用”这一事实,而不是指所有相关的文献。因为论著的作者不可能搜集所有的相关文献,同时也没有这个必要。为了准确地反映这种现象的存在程度,再导入漏引率的概念。所谓漏引率(O_r),就是指曾经引用而未列出的相关文献数占曾经引用的全部相关文献数的百分比。研究表明,在15篇遗传学发展史方面的论文应该列出719篇参考文献才能完全包括其中所含已有出处的信息内容,但实际的引文只列出216篇,占应引文献的30%,其它学科的情况与此大体相同,最一丝不苟的学者也仅列出其实际参考文献的64%^[3]。这一结果也足以说明漏引存在的严重程度。

1.3 伪引率和漏引率的量化表示

以上的研究成果表明,伪引和漏引两种现象在引文中普遍存在。为了进一步说明伪引和漏引的程度,现对其做如下量化分析:

设:

- a 表示引文中的相关文献数;
- b 表示引文中的非相关文献数;
- c 表示曾经引用而未列出的相关文献数;
- $a+b$ 表示引文中的全部文献数;
- $a+c$ 表示曾经引用的全部相关文献数。

则正常情况下可定义为:

$$P_r = \begin{cases} \frac{b}{a+b} \times 100\% & (\text{当 } a+b \neq 0 \text{ 时}) \\ 0 & (\text{当 } a+b=0, \text{ 即 } a=b=0 \text{ 时}) \end{cases}$$
$$O_r = \begin{cases} \frac{c}{a+c} \times 100\% & (\text{当 } a+c \neq 0 \text{ 时}) \\ 0 & (\text{当 } a+c=0, \text{ 即 } a=c=0 \text{ 时}) \end{cases}$$

1.4 引文、伪引和漏引关系分析

漏引的相互关系, 我们对可能产生的 8 种情形做逐

为了说明引文和曾经引用的相关文献、伪引和漏引的分析(见下表):

表 引文、曾经引用的相关文献、伪引和漏引关系分析

序号	可能发生情况	说明	引文效果	P_r O_r	现实概率
1	a=0 b=0 c=0	无引文 无曾经引用的相关文献	*	/ /	少
2	a=0 b=0 c=0	无引文 曾经引用的全部相关文献均未列出	极差	0 100%	较多
3	a=0 b=0 c=0	引文均为非相关文献 无曾经引用的相关文献	差	100% 0	极少
4	a=0 b=0 c=0	引文均为非相关文献 曾经引用的全部相关文献均未列出	极差	100% 100%	极少
5	a=0 b=0 c=0	引文中无非相关文献 曾经引用的全部相关文献均已列出	理想	0 0	少
6	a=0 b=0 c=0	引文均为相关文献 曾经引用的部分相关文献未列出	较理想	0 $\frac{c}{a+c} \times 100\%$	一般
7	a=0 b=0 c=0	引文中有非相关文献 曾经引用的全部相关文献均已列出	较理想	$\frac{b}{a+b} \times 100\%$ 0	一般
8	a=0 b=0 c=0	引文中含有相关文献和非相关文献 曾经引用的部分相关文献未列出	一般	$\frac{b}{a+b} \times 100\%$ $\frac{c}{a+c} \times 100\%$	多

注: * 分两种情况:

(1) 确实无相关文献, 其引文效果可以评价为理想, $P_r=0, O_r=0$;

(2) 如果属于未能利用现有的相关文献(不同于表中第二种情况), 多见于初写者, 是异常情况, 其引文效果应该评价为极差, $P_r=0, O_r$ 可作为 100% 处理, 这与本文 P_r 和 O_r 的定义并不矛盾。

2 伪引原因分析

(1) 引文规范方面的原因: 缺少引文规范, 作者全凭自己的意愿来确定某一论文是否作为引文列出, 造成伪引。

(2) 编辑部和审稿者方面的原因: 由于编辑部的审稿者未能对引文进行严格把关, 造成一些非相关的引文被列出。

(3) 作者方面的原因: 由于作者引文动机的多样化, 造成伪引的原因很多, 可归纳如下^[4]:

崇引。即指作者无视论文主题, 盲目引用“权

威”和“名家”的论著, 抬高自己论著的地位。这一现象无疑造成伪引。

自引。作者普遍喜欢充分引用自己的论著。据《科学引文索引》工作人员介绍, 自引率达 30%^[5]; 另对 109 位知名著者也做过统计, 自引率高达 45%^[6]。如果自引确实与论著相关倒也无可非议; 但出于炫耀或借此抬高自己的学术影响则另当别论。

转引。即指作者因某一原因而未能得到引文的原始文献, 采用该文献的中介文献加以引用的现

象。它在一定程度上也造成伪引。

策引。即指索引和自引之外,出于某种策略上的需要而有意列举一些非相关的文献,比如:故意引用多种语种的文献,故意在参考文献表中列出一大串未曾引用的文献或者引文抄袭等等。对于引文抄袭现象,荷兰科学家曾无意中发现在被引用文献在第一次引用时出现的错误,竟被后来的引用者重复多次^[7,8]。

3 漏引原因分析

(1) 引文规范方面的原因。缺少引文规范,以致作者全凭他自己的意愿来确定某一论文是否作为引文列出,造成漏引。

(2) 编辑方面的原因。主要是人为因素:

为节省篇幅,不列引文或限制引文数量。

期刊编辑不规范或有差错,如有的期刊未标明刊名、年份和卷期等信息,而作者按复印件引文,时间一长,无法再找到原文的出处,最终导致作者只好不引;有的遇到印刷出了差错,使作者找不到出处而造成错的引用。

(3) 作者方面的原因。原因很多,主要是以下几种情况:

作者(尤其是初写者)在搜集相关材料时,未能一一注明文献出处,以致有的论文无法在引文中标注出来。作者图省力,简单列几篇,未一一列出全部引文。作者对引文不重视,不管相关程度如何,随便列出其中一部分。论文的主要观点属剽窃,作者不敢标出某一或某几篇文章。作者无意中疏忽而漏掉了某些参考文献。

4 减少伪引和漏引的一些建议

为减少伪引和漏引,使伪引率和漏引率降到最低限度,笔者提出以下建议:

(1) 引文应有一个可供共同遵循的引文规范或规则。让作者列举引文时,有章可循。有的文献已对此作了一些有益的探讨^[9]。

(2) 引用的参考文献必须按行文顺序在正文中相应处进行标注,不应省略。为了做好这一工作,应尽快把参考文献标注方法列入相应的国家标准。这对期刊或其它出版物的规范化意义重大,同时对减少伪引和漏引都具有实际意义^[10]。

(3) 作者、编辑者和审稿者,三方都应齐心协力,像对待正文一样,重视引文的准确性和全面性。目前,应对引文著录格式不规范的论文,予以退回修改,不仅

要在形式上做出严格要求,还要在内容上做到引文的相关性检查。

(4) 应加强组织培训,提高引文意识,最好在大学生中开设毕业论文写作等有关课程时,能把引文的有关知识给予介绍,同时在编辑人员中开展类似的业务培训。所有的刊物应适时刊登本刊对参考文献的有关要求,并提供著录范例,以便普及和提高作者的引文知识,从而使引文分析更具有科学性、可靠性和权威性。

(5) 由于漏引现象相当普遍,美国的《艺术与人文科学引文索引》(A & H C I)对那些确实被作者利用但没有在参考文献中列出的文献也给予列出,建议《中国科学引文索引》可仿效,为解决漏引方面做些实际而又有意义的工作,为我国的文献的引文分析提供一个全面的数据来源。

参考文献

- 1 Cano, V. . Citation Behavior: Classification, Utility, and Location, JA SIS, 1989, 40(4): 284~ 290
- 2 何讯佳 引证分析中若干术语的内涵与外延之新探 情报科学, 1991, 12(4): 32~ 36
- 3 Macroberts, M. H. and Macroberts, B. R. . Problems of Citation Analysis: a Critical Review: JA SIS, 1989, 40(5): 342~ 349
- 4 宋明亮 我国图书馆学引文相关性研究 图书馆杂志, 1993, 12(3): 9~ 11
- 5 张玉华 美国科学情报所 (ISI) 考察略述 中国科技期刊研究, 1997(1): 12~ 15
- 6 邱均平 科学文献自引的统计与分析 情报学刊, 1989, 10(6): 16~ 21
- 7 Moed, H. F. and Vriens, M. . Possible Inaccuracies Occurring in Citation Analysis, Journal of Information Science, 1989, 12(2)
- 8 林蓉辉 引文分析质疑 情报科学, 1991, 12(3): 42~ 44, 49
- 9 陈益君 期刊中引文相关性的受控研究 图书情报工作, 1995(4): 18~ 21
- 10 张志琴, 熊水斌 也谈参考文献标注位置问题 编辑学报, 1997, 9(1): 46~ 48

叶新明 杭州应用工程技术学院图书馆馆员。
通讯地址: 杭州教三路 9 号。 邮编: 310012。

(来稿时间: 1997. 7. 22。 编者: 翟凤岐)