查收查引质量控制关键环节

——错引判断实践及其效果评估

□张美琦* 刘斐 姚兰 崔建华

摘要 针对普遍存在的错引现象，根据长期实践经验，研究总结人工判断错引的技巧，以把控查收查引关键环节，提升服务质量。以科学引文索引(Science Citation Index Expanded，SCIE)为例，利用案例分析法，针对科技文献引证报告服务系统提供的疑似错引信息，分析其五个数据项(作者、刊名、年份、卷/期/页码、DOI)与被引文献的匹配度，总结归纳得出错引有五类九种表现形式，把确实属于错引的引用信息纳入引用记录中。实践证明这些错引记录在增加论文影响力方面具有高度有效性。

关键词 查收查引 错引 质量控制 DOI 科研评价 效果评估

分类号 G250.7

DOI 10.16603/j.issn1002－1027.2018.05.015

1 引言

查收查引是由第三方基于文献计量学提供的一种定量评价服务，它与同行评议相辅相成，使专家能掌握足够的信息，形成依据更充分的意见，并在更高的信息集成水平上更具权威性[1]。查收查引评价的特点是：基础研究优于应用研究、整体优于个体、长期优于短期、相对指标优于绝对指标、多指标优于单个指标[2]。文献计量分析作为目前主流的研究评价方法，源于20世纪中叶兴起的科学计量学和科学引文分析[3]，20世纪80年代以来，科学计量学和科学引文分析在我国日益蓬勃发展[4－5]，查收查引也越来越受到重视，逐渐成为主流的科研评价方法之一，例如：国家杰出青年科学基金、优秀青年科学基金、国家高层次人才特殊支持计划、海外高层次人才引进计划、长江学者奖励计划等都要求检索申请者论文被主要索引数据库收录和被引用的情况，查收查引报告是科研人员或科研团队进行职称评定或报奖以及申请科研基金等的重要依据。

论文的被引用次数反映论文的影响力，是一个重要的质量评价指标。因此，引用记录的数据质量直接关系到查收查引报告的整体效率和权威性，但是，其中存在的错误引文(以下简称“错引”)信息引发了人们的重视。

错引是一种不规范的引用，之所以称之为错引，是因为该类型的引用信息与被引文献中信息存在差异，是在论文题名、作者姓名、期刊刊名、出版年份、卷期页、数字对象唯一标识符(Digital Object Unique Identifier，DOI)等数据项中至少有一处出错的引文，如：年、卷、期、页等不完全相符，但是决定文献是否相同的一些主要因素相吻合，如标题、DOI等[6]。错引的主要原因一般有两种：一种是引用者的错误。有些错引属于引用者笔误或疏忽；有些是引用者复制了他人论文的错误参考文献，属于引文失范行为[7]。另一种是数据库的录入错误。或者是施引文献信息录入错误，或者是被引文献信息录入错误。因此，在进行引用检索过程中，需要检索人员对错引信息进行判断，并将确实属于错引的引用信息纳入引用记录中，把假的错引排除掉。对于错引需要从两个角度进行把握：其一，错引也是引用记录的一部分，不能因为错，而将其漏掉或舍弃[8]。其二，疑似错引信息差别甚微，且容易引起检索人员倦怠，需要积累经验、仔细判断，不能把假的错引当成

* 通讯作者：张美琦，ORCID：0000-0001-9816-4238，邮箱：zmq@bnu.edu.cn。

真的错引,避免错上加错。

近年来对错引的研究大致可分为三类:(1)错引识别方法及成因解析,科学家群体存在引文复制等引文失范行为。1989 年,荷兰莱顿大学的莫德(Moed)和弗伦斯(Vriens)以五种期刊的错引为例,分析其分布特点及原因,发现某些作者存在"引文复制"行为[9]。2005 年加州大学洛杉矶分校的西姆金(Simkin)和罗约夫德鲁伊(Roychowdhruy)研究发现错引类型的频次分布具有指数递减规律,故推测期刊论文中 70%的参考文献来自于引文复制[10]。2007 年梁立明和钟镇以 *Nature* 上一篇高被引论文为例,探讨科学家群体中存在的引文复制等失范行为[11]。2017 年钟镇又以中国科学引文数据库(Chinese Science Citation Database,CSCD)中《中华妇产科杂志》为例,研究了中文错引识别方法及形成原因[12]。(2)错引数据的质量控制研究。1974 年,加菲尔德(Garfield)指出 SCIE 数据库收录的期刊中有错误引文信息[13],所以开发了用来纠正某些错引信息的"Keysave"系统[14]。同时,他也呼吁期刊编辑要认真对待错误引文现象[15]。2001 年苏新宁分析了编制中文社会科学引文索引(Chinese Social Sciences Citation Index,CSSCI)易出现的数据错误,并利用计算机纠正这些错误[16]。(3)错引导致引用检索时的漏检和错检研究。不同数据库都可以放宽检索条件以查到疑似错引记录,再人工判断得到确实属于错引的记录。以 SCIE 为例,错引的检索方法有两种:第一种从数据库参考文献检索入口,考虑作者、期刊的各种拼写变体,检出疑似错引记录,然后人工判断。第二种利用查收查引系统[17-18],原理与第一种方法相似,只不过系统充分考虑错引的情况,扩大检索条件检索出疑似错引记录,但是系统将相似度 90%以上的引用记录仍然提交人机交互,让人工判断真正的错引,以此来保障查全的同时具有较高的查准率。从已有成果的内容分析,人工判断真正错引相关研究的缺少,已经成为制约错引现象研究进一步展开的瓶颈之一。

2 在疑似错引记录的基础上人工判断真错引的工作实践

为了把 SCIE 数据库的疑似错引查全,"科技文献引证报告服务系统"采用的检索策略是:除了论文标题匹配,还采用"第一作者 and 刊名""第一作者 and 页码"匹配[19],根据长期的引证检索经验,这是两种引用查全率最高的检索策略,可避免数据标注不规范或者录入数据错误所导致的漏检。"第一作者 and 刊名"适用于处理期刊文献或页码缺失的文献,可有效防止错检和漏检。"第一作者 and 页码"适用于处理会议文献或页码准确的期刊文献。

查全率与查准率是一对矛盾,为保证查全率扩大检索条件,对查准率提出了严峻的挑战。如果全部结果交由机器判断,过滤条件过分严谨苛刻,必然导致漏检;过滤条件放宽,则会混入无关数据,形成错检。所以交由人工交互判断是防止漏检和错检的核心和关键。但是,人工判断主要依靠手动方式,疑似错引信息之间差别很小,劳动强度较大,导致检索人员容易出现疲劳、疏忽、倦怠等,导致错引判断准确性降低,难以应对急剧增加的引用检索服务要求,总结判断方法和技巧成为查收查引工作的迫切需求。

SCIE 数据库的每条疑似错引记录包含五个数据项:第一作者姓名、刊名、年份、卷/期/页码、DOI,但是卷/期/页码或 DOI 经常有缺失现象,根据其与被引文献题录信息的匹配度,真正错引有以下五种表现形式,详见图 1。

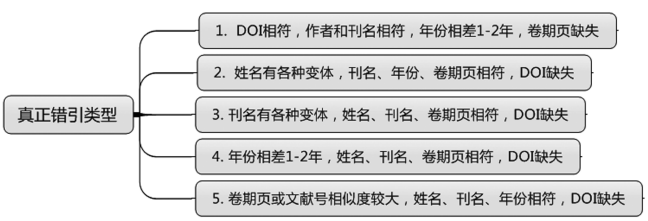

图 1 真正错引的五种表现形式

2.1 DOI 相符,作者和刊名相符,年份相差 1—2 年,卷期页缺失

DOI 是文章的唯一身份号码,只要 DOI、作者和刊名这三项相符,某作者在某期刊上发表的论文即具有唯一性。所以即使年份或卷/期/页码出现错误或缺失,也可认为是真正的错引。如例 1 所示,这两条疑似错引记录卷/期/页码缺失,但是 DOI 与被引文献相符,则认为这两条都是真正的错引。

例 1.

正确引用记录:Distinct quasi-biweekly features of the subtropical East Asian monsoon during early and late summers.

Yang,Jing;Bao,Qing;Wang,Bin;Gong,Dao-

Yi; He, Haozhe; Gao, Miao-Ni. CLIMATE DYNAMICS. 42(5－6):1469－1486. DOI:10.1007/s00382－013－1728－6.2014

疑似错引记录：

(1) Yang, J. CLIMATE DYN.():.10.1007/s00382－013－1728－6.2013

(2) Yang, J. CLIMATE DYN IN PRESS.():.10.1007/s00382－013－1728－6.2013

2.2 姓名有各种变体，刊名、年份、卷期页相符，DOI 缺失

为了把错引查全，之前放宽了检索结果的准入条件，所以引文信息中作者姓名有各种变体。以中国人姓名为例，三字姓名有 18 种变体，两字姓名有 7 种变体，以牛俊锋和杨婧的姓名为例，如表 1 所示。对作者姓名可能出现的各种拼写形式了然于胸，有助于快速准确地判断真正的错引。如例 2、例 3 所示，林国强(Lin，Guo－Qiang)的名字有 lin G 的变体，杨含心(Yang，Han－Xin)有 Han－Xin，Y 的变体。由于引用者的引用习惯不同，有的引用记录没写第一作者，只写第三作者，如例 4 所示，也认为是真正的错引。

表 1 作者姓名的各种变体

姓和名	三字姓名表达方式变体	两字姓名表达方式变体
都是全称	NiuJunFeng or JunFengNiu or Niu，JunFeng or JunFeng，Niu or Niu，Jun-Feng or Jun-Feng，Niu or Niu Jun-Feng or Jun-Feng Niu	Yang Jing or Jing Yangor Yang，Jing or Jing，Yang
一全一简	Niu，JF or Niu JF or Niu J-F or Niu，J-F or Niu J.F. or Niu，J.F. or Junfeng N or Jun-Feng，N or Niu J. or Niu，J.	Yang，J or Yang J or Yang J. or Yang，J.or Jing，Y.

例 2.

正确引用记录：Modeling and controlling the two-phase dynamics of the p53 network: a Boolean network approach.

Lin，Guo-Qiang； Ao，Bin；Chen，Jia-Wei；Wang，Wen-Xu；，Zeng-Ru. NEW JOURNAL OF PHYSICS.16():. 2014

疑似错引记录：

(1) Lin，Guo-Qiang. NEW J PHYS.16():.10.1088/1367－2630/16/12/125010.2014 (2) Lin，G. NEW J PHYS.16(12):.125010.2014

例 3.

正确引用记录：Traffic-driven epidemic outbreak on complex networks: How long does it take?

Yang，Han-Xin；Wang，Wen-Xu；Lai，Ying-Cheng. CHAOS.22(4):043146.2012

疑似错引记录：

(1) Yang，Han-Xin. CHAOS.22(4):.10.1063/1.4772967.2012

(2) Han-Xin，Y..CHAOS.22():.043146.2012

例 4.

正确引用记录：Farmers' risk preferences and their climate change adaptation strategies in the Yongqiao District，China

Jin Jianjun；Gao Yiwei；Wang Xiaomin；Pham Khanh Nam. LAND USE POLICY. 47():365－372.2015

疑似错引记录：(1) Xiaomin，w. LAND USE POLICY.47():365.2015

2.3 刊名有各种变体，姓名、刊名、卷期页相符，DOI 号缺失

为避免因刊名漏检，“科技文献引证报告服务系统”沿用了 SCIE 数据库中的刊名缩写规则和“刊名＋in press”的标注形式，采用截词符扩大刊名相似化结果等措施，扩大检索条件，提供查全率，所以要格外注意刊名的人工判断筛选。具体又分为以下四种情况：

2.3.1 作者相符，卷/期/页缺失，年份提前 1－2 年，刊名＋in press 或 inpub，说明是预印本或提前出版的情况下已被引用，如例 5 所示。

例 5.

正确引用记录：White matter pathway supporting phonological encoding in speech production: a multi-modal imaging study of brain damage patients.

Han，Zaizhu；Ma，Yujun；Gong，Gaolang；Huang，Ruiwang；Song，Luping；Bi，Yanchao. BRAIN STRUCTURE & FUNCTION.221(1):577－589.2016

疑似错引记录：(1) Han，Z. BRAIN STRUC

IN PRESS.().2014

2.3.2 作者、年份吻合相符，卷/期/页缺失，刊名＋期刊所属国家简称

如例6、例7所示，刊名后加上期刊所属国家英国的简称UK、瑞士的简称SWITZ。

例 6.

正确引用记录：Scaling behaviours in the growth of networked systems and their geometric origins

Zhang，Jiang； Li，Xintong； Wang，Xinran； Wang，Wen－Xu；Wu，Lingfei. Scientific Reports. 5 ()：. 2015.

疑似错引记录：(1) ZHANG JG.SCI REP UK. 5()：.2015

例 7.

正确引用记录：Parallel workflow tools to facilitate human brain MRI post-processing

Cui，Zaixu；Zhao，Chenxi；Gong，Gaolang. FRONTIERS IN NEUROSCIENCE.9()：.2015

疑似错引记录：(1) Cui，Zaixu. FRONT NEUROSCI － SWITZ. 9 ( ). 10. 3389/fnins. 2015. 00171.2015

2.3.3 作者、年份相符，卷/期/页缺失，刊名全称、简称、混合等各种变体写法

如例8所示，NEW JOURNAL OF PHYSICS这本期刊的变体有NEW J PHYS或J PHYS。

例 8.

正确引用记录：Exact controllability of multiplex networks.

Yuan，Zhengzhong；Zhao，Chen；Wang，Wen-Xu；Di，Zengru；Lai，Ying-Cheng. NEW JOURNAL OF PHYSICS. 16()：.2014

疑似错引记录：

(1) Yuan，Zhengzhong. NEW J PHYS. 16()：. 10.1088/1367－2630/16/10/103036.2014

(2) Yuan，Z.J PHYS.16()：.103036.2014

2.3.4 作者、年份、卷页相符，刊名写会议论文集名称或丛书名称

如例9所示，被引文献发表在2004年MICCAI (Medical Image Computing and Computer-Assisted Intervention)会议论文集中，该会议论文集又被收入LECTURE NOTES IN COMPUTER SCIENCE丛书，所以来源出版物写会议论文集或丛书，都认为是真正的错引。

例 9.

正确引用记录：Detecting functional connectivity of the cerebellum using low frequency fluctuations (LFFs)

He，Y；Zang，YF；Jiang，TZ；Liang，M；Gong，GL. MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION-MICCAI 2004，PT 2，PROCEEDINGS.3217(2)：907－915.2004

疑似错引记录：(1) He，Y. LECT NOTES COMPUT SC.3217()：907.2004

2.4 年份相差1－2年，姓名、刊名、卷/期/页相符，DOI缺失

如例10所示，年份2006与被引文献的发表年份提前1年，姓名、刊名、卷/期/页相符，则认为是真正的错引。

例 10.

正确引用记录：Effects of conservation tillage practices on winter wheat water-use efficiency and crop yield on the Loess Plateau，China

Su，Ziyou；Zhang，Jinsong；Wu，Wenliang；Cai，Dianxiong；Lv，Junjie；Jiang，Guanghui；Huang，Jian；Gao，Jun；Hartmann，Roger；Gabriels，Donald. AGRICULTURAL WATER MANAGEMENT. 87(3)：307－314. 2007.

疑似错引记录：(1) Su，Z. AGR WATER MANAGE. 87()：307.2006

2.5 卷期页或文献号相似度较大，姓名、刊名、年份相符，DOI号缺失

具体有以下两种表现形式：

2.5.1 作者、刊名、年份、卷都相符，卷期页码相似度比较大的，可以认为是笔误引发的错引

如例11所示，把“340”页，写成“(3)：40”，这本期刊本身没有期的信息，可认为是笔误。例12把“1441”写成“1141”，相似度较大，也认为笔误导致的错引。

例 11.

正确引用记录：Willingness to pay for renewable electricity：A contingent valuation study in Beijing，China. Guo，Xiurui；Liu，Haifeng；Mao，Xianqiang；Jin，Jianjun；Chen，Dongsheng；

Cheng，Shuiyuan. ENERGY POLICY.68()：340－347.2014

疑似错引记录：

(1) Guo，X.ENERG POLICY.68(3)：40.2014

例 12.

正确引用记录：Polychlorinated Biphenyls in Urban Lake Sediments from Wuhan，Central China：Occurrence，Composition，and Sedimentary Record

Yang，Zhifeng；Shen，Zhenyao；Gao，Fan；Tang，Zhenwu；Niu，Junfeng；He，Ya. JOURNAL OF ENVIRONMENTAL QUALITY. 38(4)：1441－1448.2009

疑似错引记录：

(1) Yang，Z. F. J ENVIRON QUAL. 38(4)：1141.2009

2.5.2 有些期刊没有期页，只有文献号，文献号相似度较大，也是真正的错引

如例 13 所示，文献号都是“e32766”，年份提前 1 年，也认为是真正的错引。例 14 把 013010 写成 13010，例 15 把 P05013 写成多种形式，都认为是笔误导致的错引。

例 13.

正确引用记录：Effects of Different Correlation Metrics and Preprocessing Factors on Small-World Brain Functional Networks：A Resting-State Functional MRI Study

Liang，Xia；Wang，Jinhui；Yan，Chaogan；Shu，Ni；Xu，Ke；Gong，Gaolang；He，Yong. PLOS ONE. 7(3)：. 2012. e32766. DOI：10.1371/journal.pone.0032766

疑似错引记录：

(1) Liang，Xia... Wang，Jinhui. PLOS ONE. 7(3)：.10.1371/journal.pone.0032766.2012

(2) Liang，X … Wang，J. PLOS ONE. 7()：. e32766.2012

(3) Liang，X … Wang，J.. PLOS ONE. 7()：. e32766 2011

例 14.

正确引用记录：Cooperation percolation in spatial prisoner's dilemma game

Yang，Han-Xin；Rong，Zhihai；Wang，Wen-Xu. NEW JOURNAL OF PHYSICS. 16 ()：. 2014.013010

待确认引用记录：

(1) Yang，H.-X.. NEW J PHYS. 16()：13010.2015

例 15.

正确引用记录：Cascading failure spreading on weighted heterogeneous networks

Wu，Zhi-Xi；Peng，Gang；Wang，Wen-Xu；Chan，Sammy；Wong，Eric Wing-Ming. JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT.()：.2008. P05013. DOI：10.1088/1742－5468/2008/05/P05013

疑似错引记录：

(1) Wu，Zhi-Xi. J STAT MECH-THEORY E.()：.10.1088/1742－5468/2008/05/P05013.2008

(2) Wu，Z X. J STAT MECH.()：.P05013.2008

(3) Wu，ZX. J STAT MECH-THEORY E. 2008()：050. .2008

(4) Wu，Z.－X. J STAT MECH-THEORY E. 2008(5) ：.P05013.2008

(5) Wu，Z. X. J STAT MECH THEORY E. 2008()：.050－01－050－13.2008

(6) Wu，Z. X. J STAT MECH-THEORY E.05()：.P05013.2008

(7) Wu，Z.－X. J STAT MECH-THEORY E.5()：.P05013.2008

(8) Wu，ZX. J STAT MECH-THEORY E.5()：.P05013/1－P05013/14.2008

3 错引判断的效果评估

笔者以本校 2018 年杰出青年科学基金和优秀青年科学基金申请者的 SCIE 论文数据为例，涉及环境、化学、数学、物理等学科，利用“科技文献引证报告服务系统”得到疑似错引记录，基于要素判断法和案例分析法研究人工判断真正错引的方法技巧，详见表 2。

表 2　2018 年杰出青年科学基金和优秀青年科学基金申请者 SCIE 论文错引情况

序号	作者	论文数量	错引论文数量	错引论文占总论文比率	真正错引次数	总被引次数	真正错引次数占总被引次数比率	论文起止年
1	王××	120	46	38.33%	209	4179	5.00%	2005－2017
2	董××	96	24	25.00%	91	1210	7.52%	1999－2017
3	滕××	71	17	23.94%	27	833	3.24%	2009－2017
4	龙××	68	12	17.65%	35	1703	2.06%	2005－2017
5	刘××	67	19	28.36%	78	884	8.82%	2002－2017
6	龚××	63	28	44.44%	54	2698	2.00%	2004－2017
7	苏××	59	19	32.20%	57	523	10.90%	2009－2017
8	梁××	54	15	27.78%	24	987	2.43%	2010－2017
9	毕××	53	22	41.51%	41	665	6.17%	2001－2017
10	孙××	53	9	16.98%	41	1316	3.12%	2007－2017
11	崔××	53	7	13.21%	13	739	1.76%	2010－2017
12	刘××	52	12	23.08%	22	507	4.34%	2009－2017
13	袁××	51	6	11.76%	7	848	0.83%	2008－2017
14	尹××	47	16	34.04%	27	342	7.89%	2008－2017
15	张××	44	8	18.18%	20	384	5.21%	2005－2017
16	石××	43	13	30.23%	18	437	4.12%	2004－2017
17	薛××	35	2	5.71%	2	199	1.01%	2006－2017
18	陈××	31	12	38.71%	100	434	23.04%	2012－2017
19	夏××	31	13	41.94%	70	1477	4.74%	2011－2018
20	何××	30	9	30.00%	22	332	6.63%	2011－2017
21	王××	29	13	44.83%	40	981	4.08%	2002－2017
22	郝××	28	12	42.86%	36	499	7.21%	2009－2018
23	杨××	28	12	42.86%	18	738	2.44%	2008－2017
24	杨×	27	6	22.22%	29	223	13.00%	2007－2017
25	赵××	27	8	29.63%	11	534	2.06%	2004－2017
26	蒋××	27	2	7.41%	3	276	1.09%	2009－2016
27	谢××	26	7	26.92%	15	229	6.55%	2008－2017
28	吴××	26	10	38.46%	16	249	6.43%	2006－2017
29	张×	26	8	30.77%	11	438	2.51%	2009－2017
30	王×	26	1	3.85%	1	462	0.22%	2009－2017
31	于××	24	8	33.33%	20	360	5.56%	2006－2017
32	李×	24	12	50.00%	48	908	5.29%	2009－2017
33	邢××	22	3	13.64%	3	614	0.49%	2003－2015
34	郑××	20	5	25.00%	18	307	5.86%	2012－2015
35	舒××	20	9	45.00%	15	818	1.83%	2007－2017
36	白××	19	5	26.32%	7	217	3.23%	2005－2017
37	车××	19	2	10.53%	3	338	0.89%	2009－2016
38	吴××	18	4	22.22%	20	161	12.42%	2011－2017
39	胡××	18	4	22.22%	5	83	6.02%	2006－2016
40	付××	17	8	47.06%	22	455	4.84%	2008－2016
41	刘××	17	5	29.41%	7	393	1.78%	2011－2017
42	吴××	17	4	23.53%	6	455	1.32%	2008－2009
43	包××	16	4	25.00%	5	126	3.97%	2009－2016
44	万××	16	5	31.25%	5	355	1.41%	2002－2016
45	徐××	14	5	35.71%	10	221	4.52%	2011－2017
46	潘××	14	3	21.43%	6	195	3.08%	2003－2016
47	黄××	12	3	25.00%	6	240	2.50%	2006－2016
48	毛××	11	3	27.27%	3	87	3.45%	2011－2017
49	姜××	8	3	37.50%	5	125	4.00%	2005－2017
50	卜××	7	3	42.86%	3	39	7.69%	2012－2017
平均值		34.48	9.72	28.54%	27.1	636.46	4.73%	

第一、有真正错引论文占作者论文总数的平均比率为28.54%，数据范围从3.85%－50.00%不等。从表2中可见，有真正错引论文是普遍现象，所有作者都有数量不等的论文被错引。表2作者所提交的论文中，被错引过的论文占比平均值为28.54%，接近三分之一。真正错引论文占比最高达到50%，作者一共提交了24篇论文，其中12篇论文有真正错引；真正错引论文占比最少是3.85%，作者提交的26篇论文中，有1篇有真正错引。

第二、真正错引次数占总被引次数的平均比率为4.73%，数据范围从0.22%－23.04%不等。真正错引占比最多的情况：总被引次数434次，其中有100次真正错引，占比23.04%，即五分之一还多；真正错引占比最少的情况：总被引次数462次，其中1次真正错引，占比0.22%。如果不进行人工判断，这些真正错引就被漏检了；如漏检这些真正错引，那么总被引次数的查全率必然难以保障。

总之，有真正错引论文占作者论文总数的28.54%，真正错引次数占总被引次数的4.73%，所以错引的查全率和查准率，对于量化评价有一定程度的影响。笔者所在单位自2014年至今，已经历了多次引证查询高峰期的检验，除杰出青年科学基金和优秀青年科学基金外，还有国家高层次人才特殊支持计划、海外高层次人才引进计划、长江学者奖励计划、院士申报、实验室评估、创新团队、国家奖申报等。我们根据错引的五类九种表现形式，人工判断得到真正错引，客观公正地进行量化评价，不仅增强了引证报告的权威性，而且增强了用户粘度和集体荣誉感。

4 结语

综上所述，错引是客观存在的事实，给查收查引工作带来了较大的挑战。如何控制引文质量，更加客观公正地进行科研定量评价，尚需要科学家、期刊编辑、数据库编制者和图书馆员的共同努力，笔者提出以下建议，以抛砖引玉，从而推动我国科研定量评价规范化体系的建设。

一方面，要从源头上控制引文质量，减少错引的发生。科学家从做科研伊始，就应养成严谨规范的引文习惯，减少笔误，严禁引文复制等引文失范行为；期刊编辑要认真校对参考文献信息，并利用新技术手段提高校对准确率；数据库编制者应尽最大可能提高论文数字化准确率，在数据库首页明显位置提供数据错误反馈入口，并加快审核速度，才能持续不断提高数据质量。

另一方面，图书馆员在检索时要放宽检索条件，积累错引判断方法和技巧，提高人工判断质量和效率。人工判断错引时，根据DOI、论文标题等主要因素很好判断，根据年/卷/期/页等次要因素，则要结合作者、期刊刊名、发表年份等其它数据项来综合判断，避免把假的错引当成真的错引。引文信息包含六个数据项（某些数据库导出的引文信息没有论文题名这项，所以也有五个数据项的说法）：论文题名、作者姓名、期刊刊名、年份、卷/期/页、DOI，各项都有可能出错，要有针对性地采取应对策略，具体内容如下：

（1）DOI是决定文献是否相同的主要因素之一，具有唯一性，是人工判断错引时的首选标准。但是有些论文没有DOI，例如：发表年份较早的文献没有DOI，有些期刊没有加入这个组织，其论文都没有DOI。

（2）论文题名也是决定文献是否相同的主要因素之一。该项易出现的错误是：副标题缺失、个别字词拼写错误或遗漏、合成词的连字符时有时无、英文名词有单复数变化、专业名词有简称和全称变化、希腊字母或写符号或写英文读音等。图书馆员检索时，可以只输入部分题名，不必输入全部题名，而且要剔除希腊字母、专业名词简称、数据库停用词NOT等易导致检索结果为零的字词，合成词和有单复数变化的名词能用截词符，则尽量用截词符。这样可以命中更多结果，避免漏检。人工判断时，着重核对易出现的错误点。

（3）作者姓名有各种拼写变体。以中国人姓名为例，三字姓名有18种变体，两字姓名有7种变体。姓名限定不要太单一，只要数据库支持截词符，姓名也尽量用截词符，再结合论文题名、期刊刊名、发表年份等字段进行组合检索。

（4）期刊刊名也有各种拼写变体。预印本或提前出版的情况，刊名后面会加in press或inpub标记；刊名后面也会加所属国家简称；刊名还有全称、简称、全简混合等各种变体写法；如果会议论文集又以书的形式出版，刊名可能是会议论文集名称，也可能是丛书名称，检索时也要尽量用截词符代替多种拼写变化。

(5)年份有可能相差1—2年。检索时年份字段适当往前后放宽1—2年，不要只限制在当年，否则会漏检。

(6)卷/期/页码或文献号也容易出错。作者、刊名、年份均相符，卷/期/页码相似度比较大的，可以认为是笔误引发的错引；有些期刊没有期页，只有文献号，文献号相似度较大，也可判断是笔误导致的错引。

参考文献

1 鲁索.评估科研机构的文献计量学和经济计量学指标.见：蒋国华主编.科研评价与指标国际会议论文集[M].北京：红旗出版社，2000：16—37.
2 Hicks D，Wouters P，Waltman L，et al.The Leiden Manifesto for research metrics[J].Nature，2015，520(7548)：429—431.
3 布劳温 T.科学计量学指标：32国自然科学文献与引文影响的比较分析[M].赵红州，蒋国华，译.北京：科学出版社，1989：3—4.
4 范全青，郭维真，凤元杰.我国文献计量学研究30年之发展[J].情报资料工作，2009(3)：30—33，60.
5 高俊宽.文献计量学方法在科学评价中的应用探讨[J].图书情报知识，2005(2)：14—17.
6 郝丹.引证检索中数据质量控制研究与实现[D].西安电子科技大学，2012：1—69.
7 钟镇.中文期刊与中文引文数据库错引识别方法与成因解析——以2005年《中华妇产科杂志》的CSCD错引文献为例[J].中国科技期刊研究，2017，28(03)：257—265.
8 同6：1—69.
9 Moed H F，Vriens M. Possible inaccuracies occurring in citation analysis[J]. Journal of Information Science，1989，15(2)：95—107.
10 Simkin M V，Roychowdhury V P.Stochastic modeling of citation slips[J]. Scientometrics，2005，62(3)：367—384.
11 梁立明，钟镇. 错引现象折射出的科学家群体引文失范行为——以Nature上一篇19万次高频引用论文的错引记录为例[J].自然辩证法研究，2007(6)：62—65.
12 同7：257—265.
13 Garfield E.Errors—theirs，ours and yours[J].Current Contents，1974(25)：5—6.
14 Garfield E. Project Keysave—ISI's new online system for keying citations corrects errors[J]. Current Contents，1977(7)：5—7.
15 Garfield E. Journal editors awaken to the impact of citation errors—how we control them at ISI[J]. Current Contents，1990，41：5—13.
16 苏新宁.中文社会科学引文索引(CSSCI)的设计与应用价值[J].中国图书馆学报，2012，38(5)：95—102.
17 马芳珍，李峰，季梵，刘姝，王旭，刘素清.对CALIS查收查引系统的测试和应用效果评价[J].大学图书馆学报，2016，34(2)：97—102.
18 王学勤，郝丹，郑菲，赵文忠，周津慧.“查收查引报告自动生成系统”应用实践研究[J].图书情报工作，2014，58(16)：131—137.
19 同18：131—137.

作者单位：北京师范大学图书馆，北京，100875
收稿日期：2018年7月15日

Analysis on the Key Steps of Quality Control in Citation Search Services —Practice and Evaluation of Citation Error Judgement

Zhang Meiqi　Liu Fei　Yao Lan　Cui Jianhua

Abstract：In view of the prevailing citation error，the skill of artificially judging citation error is summarized in the paper based on long-term practical experience in order to control the key steps and improve the quality of citation Search service. Using the case analysis method，the paper analyzed the matching degree between 5 data items and cited documents in SCI-expanded，such as authors，source title，publication year，Volume issue page number and DOI，based on the suspected citation error data provided by the service system. Nine types of error citation are summarized so the true citation error documents could be included in the citation counts. It has been proved that these citation error documents records are highly effective in increasing the influence of papers.

Keywords：Citation Search Services；Citation Error；Quality Control；Service System；Digital Object Unique Identifier；Research Impact Measurement；Efficiency Evaluation