

# 论文来源信息标注常见问题的分析与对策 ——以中国科学引文数据库为例

田乃庆

收稿日期 2003-07-22

修回日期 2003-12-02

中国科学院文献情报中心 100080 北京中关村北四环西路 33 号, E-mail: liannq@elong.com.cn

**摘 要** 本文通过对中国科学引文数据库 2001~2002 年收录期刊中著者地址、基金项目、页码中出现的论文来源信息标注问题,从引文数据库加工、规范、检索及文献计量学指标统计的角度进行分析探讨,并提出改进建议。

**关键词** 论文 来源信息 标注 数据库

中国科学引文数据库(以下称 CSCD)具有科技文献检索及科技活动评价功能,为我国科学文献计量和引文分析研究提供了强大工具。该库以收录期刊被引文献的检索、统计服务为特色,通过对作者及其所在机构、省市等途径,可以查询基础学科及工程技术等领域科技期刊论文的发文与被引情况,是中国科学院院士推选人和国家杰出青年基金申请项目等指定查询库。由 CSCD 核心库期刊(以下用“\*”代表)为统计源编制的《中国科学计量指标:论文与引文统计》(简称“指标集”)产品,对科学地评价我国科学活动宏观水平和微观绩效,具有重要的参考价值。

科技期刊论文来源信息包括著者地址、基金、来源期刊页码等项目,是中国科学引文数据库的重要组成部分。这些信息标注的准确与否,在一定程度上直接影响数据库加工规范,进而影响检索查询及指标集统计结果的准确性以及论文在数据库中的利用价值和社会效益。作者地址等来源信息作为最基本检索、统计依据,应严格参照国家标准 GB7713-87《科学技术报告、学位论文和学术论文的编写格式》著录,以保证数据准确、可靠。然而,目前科技论文来源信息标注的质量还不能满足引文数据库数据规范化的要求,给数据库工作造成极大困难,影响原数据的可靠性。本文以中国科学引文数据库 2001~2002 年收录期刊常见的问题为例,从加工、规范及检索、计量统计角度分析,说明论文来源信息标注质量的重要性及对数据库产生的影响,旨在为广大作者及期刊编辑校对人员提供参考,以期引起注意。

## 1 著者地址标注问题

著者地址包括作者发表论文时所处的单位、地区、省市及国别、邮编及其他联系方式。CSCD 以收录论文作者(论文任一合作者)、题名、刊名、实验室、机构、省市、关键词作为检

索入口,特别是机构、省市信息完整与否,对数据库加工、检索及统计有直接的影响。在工作实践中我们发现,作者单位及地点的著录,存在标注信息不完整、错漏较多的问题,这类错误约占所列来源信息标注问题的二分之一。

### 1.1 作者机构地点标注不全

如 2001 年《石油地球物理勘探》\*、《安徽农业科学》等刊,论文只有第一作者机构和地点,其他合著者均未注明机构所在省市,有的只注“泰探公司”等一些人们不太熟知的单位。省市地点不清楚,该字段无法加工,只能为空缺。由于地址信息在引文数据库中是必不可少的重要来源数据,因此,建议编辑部,特别对加入该库的期刊,应改进地址标注方式,以适应数据库加工、检索与统计的需要。

### 1.2 论文作者与机构对应关系出现疏漏

有的论文漏标作者单位;有的作者标注有误,作者与机构无法对应;还有的中文作者机构与英文机构标注不一致,前后矛盾,不知哪个标注正确。如:《台湾海峡》\* 2001, 20(2): 259 页 杨圣云<sup>1</sup>, 吴荔生, 陈明茹<sup>1</sup>, 周秋麟<sup>2</sup>(1. 厦门大学海洋系、亚热带海洋研究所, 福建 厦门 361005; 2. 国家海洋局第三海洋研究所, 福建 厦门 361005)。第二位作者漏注单位标号,无法录入机构名称。

### 1.3 机构中的国家和重点开放实验室名称不规范

作者机构中出现的实验室名称,应按照其正式、完整的名称标注,同时应注明以下所属类别:国家重点实验室和部门开放、重点实验室,以及省、市、局、公司重点实验室、国家工程研究中心名称。在规范数据时我们发现,同一个实验室有多种写法,如“人工微结构物理与介观物理国家重点实验室”,有的注“人工介观物理”、“介观物理”、“人工物理与介观物理”等名称达七种之多。像“分子生物学重点实验室”等名称近似或一字之差甚至同一名称实验室较多,各部委、中国科技期刊研究, 2004, 15(4)

省市、国家重点实验室等均有类似名称,称谓不准确或没注明归属部门,非常容易混淆,应特别明确标注。

上述著者地址标注问题对数据库的影响:

### (1)对数据加工、规范处理的影响

作者机构、地点是数据库加工必备的字段,对于引文库数据规范时作者、机构之间的链接、数据改错、判断具有参考价值。如果信息不准确或作者与其机构缺乏对应关系,就无法完成作者与机构、省市字段的加工、定位,也就不能发挥其应有的作用。

### (2)对用户检索的影响

作者机构、省市作为信息要素之一,在情报检索中充当重要的角色<sup>[1]</sup>。利用机构、实验室、省市字段,可对特定需要增加其检索的方式及途径,尤其对于同名同姓不同学科作者发文及被引的检索,增加限定条件的控制,起到重要的排重、鉴别作用,大大提高查全率和查准率,保证检索结果的详实性。因此,作者机构、省市数据的缺陷,将使检索不能获得预期效果。

### (3)对计量指标统计的影响

作者发文机构、地域与引文数据的链接,对统计所在机构、地区论文产出力和影响力研究指标,反映某机构、地域在某学科的科学研究状况和水平具有重要作用。而著者地址标注不全和不确定,此项工作就难以客观准确地进行。特别对于国际、国内合作研究指标的统计,当以全部作者为统计源时,其机构、地点数据空缺,将不能进入统计。另外,在实验室统计中,实验室所在地点对实验室名称的规范、合并具有一定的参考作用。例如,在统计发表 CSCD 论文最多的前 10 名实验室时,“教育部胶体与界面化学开放实验室”论文为 39 篇;“胶体与界面化学开放实验室”3 篇,另有“胶体与界面科学开放实验室”7 篇,前二者地点为山东,后者为北京,这时根据实验室所在地的不同,可以准确判断前两个同为教育部实验室,而后者是中国科学院实验室。机构地点对排除实验室名称有误造成的混淆极为有用,如果没有标注地点,易造成判断失误,使统计名次过高或榜上无名,影响结果的准确性。

### (4)对作者机构、实验室显示度的影响

从计量统计角度,作者与所在机构、实验室、地点有一一对应关系,有姓名“无单位”,有机构无“地址”及实验室名称的不准确等情况,无疑影响作者机构、地区和实验室的发文(被引)统计频次。比如,以 2001 年 CSCD 发表国家自然科学基金论文最多的前 10 名实验室统计为例,其中“环境化学国家重点实验室”进入前 10 名之列,论文数量为 45 篇,而该实验室写成“水环境化学国家重点实验室”论文数量 3 篇,由于实验室名称按字顺排序,极易造成这 3 篇数据的遗漏,影响统计名次。因此,实验室名称应按规范名称标注,不可

随意而写。

机构、地点的标注及国家重点实验室和部门开放实验室名称的规范写法对于作者是轻而易举的事情,但对于数据库工作人员,为保证检索及统计准确性,不仅要对各种名称做规范整理,还要对作者与单位、地点不挂钩的情况进行处理及对有疑问数据的核查,需要花费很大精力。因此,为减少不必要的麻烦,提高准确率,论文原始信息一定要准确、可靠。

## 2 基金标注问题

我国现行的基金制度,对科学研究的发展和科技创新事业起到很大推动作用。基金论文的标注是基金管理制度的体现,基金论文的产出是衡量期刊学术质量的重要评价指标<sup>[2]</sup>。因此,基金的标注对于研究基金论文的机构、学科、地区分布研究指标的统计具有重要意义,这就要求基金标注内容与形式应准确完整。

正确的基金标注应反映出基金资助的所属部门及基金名称。但在数据规范中,我们发现,基金名称标注存在不少问题,主要有以下几方面。

### 2.1 基金项目省略所属机构

省略所属省、市、校等名称。这种情况较多,特别是两个基金合写时更常见。例:《兵工学报》\* 2001,22(4):115 页:2000 年度校科研基金。由于 CSCD 在做基金规范时,将基金项目的数据单独拆分为基金规范库,与原数据的机构字段脱离。因此,在著录各省市机构的基金名称时,不应省略其省市、机构的所属名称,这样造成基金数据的原省市、机构名称脱节,给基金数据的规范造成查找、挂靠上的麻烦,也易使数据出错。

还有如:“人才培养基金”、“千百十人才工程基金”、“青年骨干基金”等等,未注明资助来源,许多省市、部委及所属的系统都有该项基金,无法做出准确判断。

### 2.2 基金名称用英文首字母缩写代替

英文版期刊的标注,此类情况较多。如:《Acta Mathematica Science》\* 2001,21(1):137 页:ZARCF;同刊,2002,22(4)517 页:PNSFS。另外,类似“TY of China”等也不易准确译成中文名称。

### 2.3 基金名称不准确

缺少具体名称。如《化学学报》\* 2001,59(1):6 页及该刊(9):1384 页:西南石油学院国家重点实验室基金资助。由于高校的科研实力较强,一般有多个国家重点、开放实验室,因此应著录清楚,否则,不易分辨。

基金项目及国家重点实验室资助名称应该完整,不应省略全称或简写,以免在规范基金名称时不易识别。实验室资  
中国科技期刊研究,2004,15(4)

助项目,可参照机构中实验室的标注要求,名称应完整、规范。上述基金标注作法不可取,容易造成数据分散而减少统计数量,尤其缺少所属部门的基金,无法将其按应所在的类别统计,从而造成数据的损失,最终影响基金统计结果的准确性。应当指出,作者对自己特定领域基金的准确名称及名称所指是清楚的,但对于数据库工作人员,面对如此大量五花八门的数据,在规范中由于查找资料及了解途径的局限,要使基金数据规范完全准确、客观是有一定难度的。所以,作者应尽可能提供完整的原始数据,只有这样,才能使基金统计更准确、客观。

### 3 来源期刊的页码问题

#### 3.1 装订丢页、重页 造成某些页码混乱

有的期刊一期内就有多处页码装订错误,致使数据残缺不全。如《中国药物依赖性杂志》2001,10(4):缺少261-262,267,268页,重复259-260,269-270,致使三篇论文不完整:丢失两篇来源页及一篇引文页和一页正文,无法正常加工。

#### 3.2 页码标识、编排错误

转页页码标注有误,如《压电与声光》\*2001,23(4):306-308转312页,在转页上应注“上接308页”,而错写为“上接289页”,造成衔接错误。还有页码编排的疏漏,如《红外与激光工程》2001第4期页码:从“157-231”页,正好与该刊第3期页码重合,正确页码应编为“252-309”。

期刊论文具有为读者参考、利用的价值。上述来源期刊页码装订、排版错误,使刊物不能提供准确的论文页码及内容,因而无法进行正常数据加工,也给读者查阅、检索带来困难,必定影响该文献的引用。论文信息的完整性,是期刊整体质量的起码要求。如果论文学术水平再高,但由于缺页使读者无法利用,将不能体现其应有价值,期刊质量形象也会受到影响。因此,来源期刊的页码,作为来源数据基础要素和日后引用参考文献的依据,必须准确无误。

### 4 论文标注问题产生的原因及对策

#### 4.1 重正文轻辅文

论文主体是学术水平的体现,而其辅文部分标注是论文传播、利用、服务的工具,是不可忽视的论文有机组成部分。从上述分析可知,来源信息具有检索、识别作用,各字段之间存在相互联系,特别是作者、机构及地点的标注,在计量指标统计中扮演重要角色。因此,编校人员不仅要关注论文主

体,还应重视对论文信息标注的仔细检查、校对。信息标注出现的一些错漏,反映了一些作者及编校人员对论文标注内容重视程度不够。对于作者与机构对应关系、页码标识等的疏忽,只要作者和编校人员再认真、细致一些,是完全可以避免的。

#### 4.2 责任意识与品牌意识淡薄

科技论文发表后,除了为广大读者提供传播信息,还为计算机检索、他人引用和论文统计提供依据,对计量统计和期刊评价产生影响,这意味着一种社会责任。如果期刊来源信息标注错误,还会引起二次文献加工及其他载体的循环错误,使读者很难追溯原本正确的信息。因此,应从一次文献严格把关。作为作者,提供的信息一定要准确;作为入编数据库的编校人员,特别是入选CSCD核心库期刊的编校人员,应把信息标注完整与准确,同期刊整体质量、期刊的声誉和编校人员工作水平紧密相联。建议有条件的编辑部做好刊后审读工作,查疑补漏。这是期刊自身总结经验和自我完善的一项行之有效的办法,是期刊质量的重要保证<sup>[3]</sup>。

#### 4.3 国家标准有关规定不够明确

笔者注意到,国标中有关作者地址标注,没有作特别要求。如GB7713-87《科学技术报告、学位论文和学术论文的编写格式》-5.1.2.g中对论文前置部分责任者姓名描述:“必要时可注明个人责任者的职务、职称、学位、所在单位名称及地址”。由此,可能造成一些期刊作者地址标注不够详细。随着社会信息量日益增加和社会需求的不断加大,数据库得到迅速发展。因此,为适应当今形势变化和数据库加工、检索需要,建议有关部门,对国家标准中作者单位及所在地址作明确规定,制定著录细则,使地址信息规范化,以满足人们获取文献信息的需要。

综上所述,论文信息来源标注对数据库的加工、检索及计量指标的统计影响不可低估,对期刊整体水平和功能的发挥具有重要的作用。只要广大作者和编校人员有足够的重视,科技论文在数据库将体现其应有地位和价值,产生重要的社会效益。

#### 参考文献

- 何建新,胡方方.论著者地址的情报价值.编辑学报,1999,11(1):13~14
- 张英健.部分高校自科版学报基金论文刊载情况的研究.盐城工学院学报,2002,15(4):78~80
- 颜巧元,王菊香.刊后集体审读有利于提高期刊质量.编辑学报,2002,14(1):61