

论被引频次评价的适应性、局限性和不合理性

蔡言厚, 杨 华

(中南大学 高等教育研究所, 湖南 长沙 410083)

摘要:被引频次是文献计量学中被用来测度学术论文社会显示度和学术影响力的重要指标。被引频次评价是指用学术论文发表以后被引用的次数来评价以学术论文形态表征的研究成果,是评价与学术论文有关的期刊、学科(专业)、国家(地区)、单位(个人)的一种方法。被引频次评价是由美国科学信息研究所倡导的、国际上广泛公认的、以学术论文为载体的研究成果评价体系。目前,被引频次评价似乎已经成为评价期刊学术论文质量和学科(专业)、国家(地区)、单位(个人)学术水平的唯一方法。其实,被引频次评价不是万能的,它有一定的适用范围。被引频次评价只适应于基础研究,不适应于开发(发展)研究,在应用研究方面有局限性。总之,被引频次评价有它的适应性、局限性和不合理性。

关键词:被引频次;评价;适应性;局限性;不合理性

中图分类号:C1 **文献标志码:**A **文章编号:**1008-5831(2009)05-0059-04

被引频次评价是指用学术论文发表以后被引用的次数来评价以学术论文形态表征的研究成果,是评价与学术论文有关的期刊、学科(专业)、国家(地区)、单位(个人)的一种方法。而在现实中被引频次评价似乎已经成为评价期刊学术论文质量和学科(专业)、国家(地区)、单位(个人)学术水平的唯一方法。其实,被引频次评价不是万能的,它有一定的实用范围。在基础研究、应用研究和开发(发展)研究的成果评价中,被引频次评价只适应于基础研究,不适应于开发(发展)研究,在应用研究方面也有局限性。这一点在国科发基础字[2003]308号文《科学技术评价办法》(试行)中已有清楚表述。在此背景下,笔者仅对被引频次评价的适应性、局限性和不合理性进行分析。

一、被引频次评价的适应性

目前,国际上在判断一篇学术论文的价值时,通常用该学术论文发表以后的被引频次来评价。学术论文的被引用次数越多,说明该学术论文在同行中引起的反响越大,受同行关注的程度越高。也就是说,被引频次评价是评价与学术论文有关的期刊、学科(专业)、国家(地区)、单位(个人)的一种暂时无法替代的有效方法。因为被引频次评价是通过数据库来实现的;而用于被引频次评价的数据库选刊都是按尤金·加菲尔德(Eugene Garfield)的浓缩理论来进行的。Garfield引用了著名的“80/20规则”,认为精选20%有代表性的刊物,可以有80%的有用信息量。只要选刊正确,可以做到以部分代全体^[1]。因此,被引频次评价适应于基础研究成果。不仅如此,实践证明:被引频次评价还适用于同类期刊和学科(专业)以及同一学科(专业)内的国家(地区)、单位(个人)学术研究评价。

收稿日期:2009-05-31

作者简介:蔡言厚(1937-),男,中南大学高等教育研究所教授,主要从事高等教育及评价研究。

被引频次评价的适应性取决于数据库,不同的数据库适应于不同的评价对象。目前全世界出版的现期期刊约20万种,其中科技期刊约7万种,人文社科期刊约3万种,其他一般性、通俗性、时政性、消遣性期刊约10万种。面对如此庞大的期刊数量,任何一个期刊数据库都不可能也不需要收录所有的期刊。实际上,各种数据库都是根据其编制宗旨、用户信息需求来选择一定的期刊作为收录的对象。目的越明确,定位越准确,被收录期刊的选择就越严格^[2]。

目前,中国一些评价机构在运用单项指标或多项指标对单位进行评价时常采用的数据库有:ESI(基本科学指标)、SCI(科学引文索引)、SSCI(社会科学引文索引)、EI(工程索引)、ISTP(科技会议录索引)、ISSHP(社会科学及人文科学会议录索引)、A&HCI(艺术和人文科学引文索引)、CSTPCD(中国科技论文与引文数据库)、CSSCI(中文社会科学引文索引)、CSCD(中国科学引文数据库)、CNKI(中国引文数据库)以及《中国大学评价》课题组研制的SCD。

显然,用同一数据库的学术论文被引频次来评价该数据库所收录期刊的学术水平时,一般来说是可行的。但是,要进行客观、公正的评价,最好是将被收录期刊进行分类评价。如果不进行分类评价,在同一数据库中,对反映基础研究成果的期刊和反映应用技术成果的期刊进行统一排名时,就不能做到客观和公正。

用同一数据库的学术论文被引频次,与用不同数据库的学术论文被引频次对它们所涵盖的各个单位的学科(专业)进行评价和排名,都是比较客观和公正的。因为学科(专业)本身就具有类属特征,各个单位的学科(专业)在同一的还是不同的数据库中所收录的学术论文被引频次机会均等。

用同一数据库的学术论文被引频次,与用不同数据库的学术论文被引频次对它们所涵盖的各个学科(专业)内的国家(地区)或单位(个人)进行评价和排名都是可行的。因为在经济全球化的今天,一个国家(地区)或一个单位(个人),无论大小,无论在哪里,如果在科学技术上没有特长,是不能生存下去的。只有拥有在世界范围内领先的科学技术,才能立于世界民族之林。这样做将有利于各个国家(地区)或单位(个人)参与世界竞争,从而促进世界的发展。

综上所述,一个数据库只要遵循“80/20规则”选刊,被引频次不仅适应于基础研究成果的评价,还可适用于同类期刊和学科(专业)以及同一学科(专业)内的国家(地区)、单位(个人)的评价。

二、被引频次评价的局限性

由于应用研究成果是以应用科学技术知识在科学研究、技术开发、后续开发和应用推广中取得新技术、新产品,获得自主知识产权,促进生产力水平提高,实现经济和社会效益为评价重点,因此,应用技

术成果的技术指标、投入产出比和潜在市场经济价值等应作为评价的重要参考指标。所以,用被引频次评价应用研究成果时存在局限性。

被引频次和学术论文排名一样,被引频次评价的局限性和学术论文排名的局限性也大致相同。它表现在数据库使用的“包装”、用一个数据库的学术论文及其被引频次来评价不同的对象和使用多个数据库的学术论文及其被引频次来评价同一对象等多个方面。

中国在20世纪80年代曾经用SCI(科学引文索引)所收录的学术论文对中国(含台湾)大学进行排名;20世纪90年代使用SCI所收录的学术论文对中国和世界部分大学进行排名;近几年有的使用ESI(基本科学指标)对世界大学科研竞争力进行排名,有的使用SCI、EI、SSCI、CSSCI收录的学术论文对大学的学术成果进行排名。如果用某数据库的学术论文对同一类型的大学或研究机构进行排名是无可非议的;同样用某数据库的学术论文被引频次对同一类型的大学或研究机构进行排名也同样具有合理性;用某数据库的学术论文和被引频次对大学或研究机构进行排名,并如实标明数据库名称是可以接受的。

但是,用SCI所收录的学术论文对中国(含台湾)前30所大学进行排名的主办者是用一个数据库的学术论文来评价不同类型的大学,所得到的排名结果令人啼笑皆非:清华大学在中国(含台湾)前30所大学排名中排在17位^[3],排在清华大学前面的大陆大学,除北京大学外,还有其它11所大学。更令人难以接受的是发布的题目:学术榜——中国科学计量指标排序。这里犯了两个常识性的错误:一是中国大陆大学是按行政的意志分综合、理、工、农、医、师范等类型办学的,二是不能用一个标准去衡量不同的事物。众所周知,20世纪50年代中国按照苏联模式对高等教育进行了大规模的调整,使清华大学由一个综合性大学变成典型的工科院校,而SCI所收录的学术论文是偏重基础学科和基础研究的,这样的评价体系对清华大学非常不利。也就是说用一个标准去衡量不同的事物,显然是行不通的。如果实事求是地标明用《SCI(科学引文索引)》收录的学术论文排名,那也可以接受;问题在于发布时进行了“包装”,使SCI(科学引文索引)所收录的学术论文排名变成了“学术榜”,这就使人难以接受:难道清华大学的学术水平还不如北京大学以外的那些学校?“学术”包括科学和技术,如果说当时清华大学的科学水平不如那些大学中的某个学校,还说得过去;那么清华大学的技术水平绝对是大陆大学第一。经过广泛听取意见后,主办者于1989年分别用EI(工程索引)和SCI(科学引文索引)所收录的学术论文对中国(含台湾)前20所工科大学和前20所理科大学进行排名,得到了一个比较满意的结果^[4]:清华大学在中国前20所工科大学科学计量单项指标排序中排第3名(在大陆排第1),北京大学在中国前20所理科大学科学计量单项指标排序中排第2名

(在大陆排第1)。

20世纪90年代使用SCI所收录的学术论文对中国和世界部分大学进行排名的主办者犯的是同样的错误。

使用ESI(基本科学指标)对世界大学科研竞争力进行排名,使用SCI、EI、SSCI、CSSCI收录的学术论文对大学的学术成果进行排名,都有包装的嫌疑。

使用ESI对世界大学科研竞争力进行排名的主办者已经意识到这一点。在2007年再用ESI对世界大学科研竞争力进行排名时,增加了DII(专利数据库)数据。但是,只凭这两项指标是不能表征大学竞争力的,大学竞争力应包括大学的学科资源、财物资源、人才资源、校园文化、社会声誉等,并在教学、科研、社会服务某个或某些方面显现出比其他大学强的能力,这些能力只用ESI和DI将难以涵盖。

使用SCI、EI、SSCI、CSSCI收录的学术论文对大学的学术成果进行排名除有包装的嫌疑外,更主要的是反映主办者的大学评价宗旨存在问题。主办者不是不懂得学术成果的含义,而是由其大学评价宗旨决定的。主办者的大学评价宗旨是以从事基础研究为标杆而选择评价指标。众所周知,大学进行的科学研究所形成的研究成果包括学术论文、专利和非专利技术形成的新技术、新产品。基础学科的科学研究成果主要形态是学术论文,而应用学科、工程技术学科的科学研究成果主要形态是专利和非专利技术形成的新技术、新产品。这里所犯的错误仍然是用一个标准去衡量不同的事物。

三、被引频次评价的不合理性

目前,社会上比较公认的三大综合排行榜的主办者之一,面对社会的质疑,对使用S&N、A&HCI、EI、SSCI收录学术论文和专著,SCI和CSSCI、CSCD、SCD收录论文及其引用频次,授权专利、获奖国家级成果对大学的科学研究和研究生院进行排名,不断改进和完善,主要通过调整三级指标的权重来实现。主观愿望是一回事,客观效果又是一回事。调整三级指标权重的结果是使科学研究和研究生院的排行榜变成学术论文排行榜^[5]。文献[5]的研究结果充分证明了这一点。文献[5]以自然科学为例,根据主办者历年提供三级指标的权重和相应年度的有关数据,得到1997年的论文类总得分、研究成果类总得分、专利类总得分在自然科学研究总得分中的比例分别是:52%,45%,3%;2002年,论文类总得分、研究成果类总得分、专利类总得分在自然科学研究总得分中的比例分别是:86%,13%,3%;而2006年,论文类总得分、研究成果类总得分、专利类总得分在自然科学研究总得分中的比例分别是:84%,2%,14%。也就是说,研究成果类总得分、专利类总得分无足轻重。自然科学是这样,社会科学也是这样。实际上使用S&N、A&HCI、EI、SSCI收录学术论文和专著,SCI和CSSCI、CSCD、SCD收录论文及其引用频次,授权专利、获奖国家级成果对大学的科学研究和研究生院进行排名就是学术论文排行榜。2008年,主办者想通过提高引用频次的权重来提高排行

榜的质量,这一措施一般有利于以基础学科见长的大学,使这些大学名次上升,而使以工科为主体的大学名次下降。如中国科技大学的名次得到很大的提高,从2006年的11名提升到2008年的第7名,使华中科技大学由第7名退至第8名,使南开大学与天津大学、北京师范大学与中南大学易位^[6]。用学术论文来对所有的大学进行排名是不合理的,而采用提高引用频次权重的措施提高排行榜的质量,只对以基础学科为主体的大学是合理的,对以工科为主体的大学是不合理的,对以工程学科、竞技学科、实践性学科为主体的院校更不合理。因此,对三级指标权重进行修改,主要是提高研究成果类——国家级一等奖、国家级二等奖的权重。其具体做法就是保持论文类和专利类指标权重不变,只对研究成果类指标的权重进行修改。修改要遵循非德尔法原则、统计原则和导向原则^[5]。

所谓遵循非德尔法原则,就是修改指标权重时应该以1997年1927名专家连续三轮共4448份书面意见的中位数确定的论文、研究成果、专利三类指标的权重分配数值比例为基准^[7]。即在同类型的指标中,把取消的指标权重加到保留的指标上,增加的指标权重从原来的同类指标中分得。例如,研究成果:1997年时由自然科学的“四个类别鉴定成果、三个类别奖励成果”和社会科学的“向有关部门提交的成果、鉴定成果、三个类别的奖励成果”,而到2000年时取消了研究成果类别中的“所有的鉴定成果”、“地市级奖励成果”、“向有关部门提交的成果”,只保留“国家级奖励成果”和“省部级奖励成果”;到2003年时又取消了研究成果类别中的“省部级奖励成果”,只剩下“国家级奖励成果”——“国家级一等奖”和“国家级二等奖”。在修改三级指标权重时应该把自然科学的“四个类别鉴定成果、另两个类别奖励成果”和社会科学的“向有关部门提交的成果、鉴定成果、另两个类别的奖励成果”的权重加到“国家级奖励成果”——“国家级一等奖”和“国家级二等奖”的权重上。

所谓统计原则,就是修改指标权重时要对当年的高等学校拥有的各类指标总量进行统计,按照当年三级指标的权重计算出高等学校拥有的各类指标总得分,并与1997年各类指标总得分进行比较,其得分比例是否一致。如果一致,那就不用修改;如果不一致,就要对三级指标进行修改,其具体做法就是设计多组数据试算,找出接近1997年各类指标总得分比例的一组作为修改的依据,并进行微调,尽可能与1997年的各类指标总得分比例相一致。

所谓导向原则,就是根据国家对于大学的要求,引导各个大学向国家要求的方向发展,积极投入到国家急需发展的事业中。例如,中国要实施创新型国家建设,就需要各个大学办成创新型大学,鼓励各个大学在拥有自主知识产权方面加大力度,进行原创性研究,在“专利类”成果方面应提高其权重。

四、结论

通过以上讨论,可以得到如下三点结论。

其一,被引频次与学术论文一样,只适应于基础研究成果评价和同类期刊、学科(专业)以及同一学科(专业)内的国家(地区)、单位(个人)的评价。

其二,被引频次评价使用在应用研究时是有局限性的,表现在数据库使用的“包装”、用一个数据库的学术论文及其被引频次来评价不同的对象和使用多个数据库的学术论文及其被引频次来评价同一对象。

其三,提高引用频次的权重来提高排行榜的质量,导致研究生院的排行榜变成学术论文排行榜;用学术论文来对所有的大学进行排名是不合理的,而采用提高引用频次权重的措施提高排行榜的质量,只对以基础学科为主体的大学是合理的,对以工科为主体的大学是不合理的,对以工程学科、竞技学科、实践性学科为主体的院校更不合理。

参考文献:

- [1]师昌绪,田中卓,黄孝瑛,等. 什么是科学引文索引(SCI)及我国所处的学术地位[J]. 自然科学进展,1997(4): 506-512.
- [2]叶继元. 引文数据库精选来源期刊对学术评价作用的分析[J]. 云梦学刊,2004(4):17-19
- [3]中国管理科学研究院科学学所. 学术榜——我国科学计量指标排序[N]. 科技日报,1987-09-13.
- [4]中国管理科学研究院科学学所. 学术榜——我国科学计量指标排序[J]. 瞭望,1989(2): 27.
- [5]蔡言厚,蔡莉.“中国大学评价”三级指标权重需要修改[J]. 中国高等教育评估,2007(2). 30-34.
- [6]武书连. 挑大学 选专业[M]. 北京:中国统计出版社,2008:1,58.
- [7]武书连,吕嘉,郭石林. 中国大学研究与发展成果评价(节录)[J]. 科学学与科学技术管理,1997(7):49-72

On the Adaptability, Limitation and Irrationality of Appraisal by Frequency of Quotation

CAI Yan-hou, YANG Hua

(Institute of Higher Education, Zhongnan University, Changsha 410083, Chian)

Abstract: Frequency of quotation is an important index to measure the social prominence and academic influence of a paper in document metrology. Through measuring the times being quoted after the paper is published, appraisal by frequency of quotation evaluates research achievement in the form of academic papers. It is a way to evaluate periodicals, disciplines (majors), countries (districts) and units (individuals) related to certain academic papers. Appraisal by frequency of quotation is a research achievement appraisal system with academic papers as its carrier advocated by American Scientific Information Research Institute and widely acknowledged in the world. Nowadays it seems to be the only method to appraise the quality of papers published in journals and the academic standard of a discipline (major), a country (district) and a unit (individual). Actually, appraisal by frequency of quotation is not omnipotent; it has its own range of application. It can be applied to basic research but not to development research. It has its limitation when applied to application research. In one word, appraisal by frequency of quotation has its adaptability, limitation and irrationality.

Key words: frequency of quotation; appraisal; adaptability; limitation; irrationality

(责任编辑 彭建国)