

不要再玩数字游戏 *

关键词：学术评价

作者：大卫·罗格·班纳时 (David Lorge Parnas)

译者：张 健 朱雪阳

数论文式评价延缓了科研进程。

译者按：长期以来，国内对科研成果进行量化评价的做法普遍存在，但是人们已经逐渐认识到这种评价方式对科研进步的阻碍。2007年，软件界先驱大卫·班纳时 (David Parnas) 教授发表的这篇关于科研评价量化问题的文章，针砭时弊，鞭策入理，时隔多年再读，仍如醍醐灌顶。CCF 公共政策委员会决定翻译此文以飨读者。在翻译过程中，我们多次与原文作者及其夫人植丽颜 (Lillian) 女士讨论。作者特别谈到，在原文刊出后，多位其他领域、其他语种的科研人员对此文颇感认同。在当前破“五唯”的背景下，希望此文为我国计算机领域科研评价工作提供借鉴。

作为一名资深的科研人员，我很痛心地看着，科研资助机构、系主任、学院院长、职称评审委员会鼓励年轻的科研人员去做浅显的研究。作为学术期刊的读者（这些期刊本应该是严肃的研究性期刊），我很恼火地看到，计算机学术界被越来越多的科学价值越来越小的文章污染了。作为期刊编辑或审稿人，当我听到一些暗示期刊要为作者（而不是读者）服务的讨论时，我就很生气。学术期刊的其他读者理应像我一样感到愤怒，并要求改变这种状况。

引起这些现象的原因是这种被普遍采用的政策：用科研人员发表论文的数量来评价他们，而不是他们学术贡献的正确性、重要性、创新性、相关性。这种只是简单地数论文而不是阅读并评估论文的普遍做法，根本就是错误的，原因如下：

它鼓励浅薄的研究。那些发表了很多仓促写成的、肤浅的（甚至是错误的）论文的科研人员，其排名将比那些花费多年时间研究重要问题的科研人

员更靠前。换句话说，数论文的做法只是统计数量，而不是质量或价值。

它鼓励过分庞大的研究团队。那些领导大团队的学者，往往很少花时间与学生单独讨论，但在所有学生的论文上都进行署名，其排名将比领导小团队而与其成员紧密合作的学者更靠前。

它鼓励重复性工作。有些科研人员采用“拷贝 - 粘贴 - 伪装”模式在诸多会议和期刊上发表相同的工作，他们比那些只发表新想法或新结果的学者可获得更高的分值。

它鼓励小的、无关紧要的研究。有的科研人员只根据三四个学生的简单观察就发表所谓的“经验性研究”，其排名比那些从事长期、复杂实验研究的科研人员更靠前。

它奖励发表半成品的做法。有的科研人员只是描述某种语言和系统，但并不实现也不使用它们，其排名却比那些真正进行实现和实验的科研人员更靠前。

* 本文译自 *Communications of the ACM*, “Stop the Numbers Game”, 2007, 50(11): 19~21 一文。

往往有人会辩解说,基于数论文的评价方案是“客观”的,它比仔细阅读论文的方式更省时,成本更低。遗憾的是,这种“客观的”对学术贡献的评价方法经常与学术贡献无关。

数论文评价方式的支持者认为,只有好论文才能发表于“最好的”期刊;对这些论文,没必要再去读了。但实际上,做过编委的人都知道,在评审过程中,不同的评审人在严肃性、客观性、细致性等方面有很大的差异。他们经常互相矛盾,也可能会犯错误。很多编委并不想去深入探究并解决问题;他们只是简单地算一个平均分,然后把评审意见转给作者。被一个会议或期刊拒绝的文章,往往(不用修改就)被另一个会议或期刊接受。偶尔,开始被拒收的论文后来却获奖了;而有些被接受的论文到头来却是有错的。即使是细心的评审人和编委,一次也只评审(一位作者的)一篇文章;他们可能并不知道作者以其他题目和摘要发表过这一工作。如果你信任这样的评审过程,那是愚蠢的。

以发表论文的数量来衡量科研人员的工作效率,会延缓科研进程。为了提高得分,科研人员肯定不会考虑去解决那些难题,也不会考虑去解决那些可能需要经年专注工作才有望解决的问题,而只会去研究一些较容易的问题。

基于发表论文数量的评估可能会腐蚀科研人员,让他们学会“上有政策下有对策”。如果他们知道只有论文数量管用,可能会采取如下对策:

签订发表论文的协议。“如果你的文章署我的名字,我也在我的文章中署你的名字”。如果4~6个科研人员作为团队一起采用这样的策略,非常有效。有时,我会遇到据称是某篇论文的“作者”但此人实际上并没有读过那篇文章的情况。

组团。一些研究人员组成小团体(派系),用一些特别的术语讨论一个狭窄的话题,这些话题的广度仅能达到一个系列会议或一个期刊的基本要求。然后,他们就“从派系到派系”发表论文。组织这样的派系,虽然会提高派系成员的产出,但是它会导致劣质的沟通以及重复性工作,因而不利于科研工作的推进。

随意发表(怎么都行)。有些科研人员尽管知道自己的文章可能有错、过时、不切题,但仍然会去发表。因为他们知道,只要文章能通过评审就算数。

定制式研究。有些科研人员时刻注意各种会议和期刊专辑的信息,针对其征文通知,提交定制好论文(可能是由事先准备好的论文片段加工而得)。

可发表的最小增值单元(Minimum Publishable Increment, MPI)。很多科研人员在完成了一个深刻的研究之后,对成果进行切分,以发表尽可能多的论文。每篇论文只包含刚刚够发表的新内容,但是其研究动机和背景知识与之前的其他论文重复。在发表了这些小论文(MPI)之后,作者再发表一篇综述(major review)。其实,如果只发表一篇文章,更有利于推进科研。

组织学术会议和小型研讨会。有些科研人员发起一些专门的学术会议和小型研讨会。这些会议的论文非常集中,而且几乎肯定会发表会议组织者的论文。会议结束后,会议论文集再加上“前言”,作为一本书正式出版。这样的做法,使得会议组织者能多发表三篇文章:会议论文、书的章节、前言。

参加会议时,你就可以看到这些游戏的结果:人们参会只为作报告,而非倾听;报告厅里经常是听众寥寥,有些参会者根本就不去听报告。

为了改善这一明显的缺陷,一些评估者把引文数加入评估系统。在这里,未读论文仍是一个致命错误。因为有些引用是否定相关研究的,还有一些引用只为了说明其他人对该主题感兴趣,或者说明作者知道这些文献。有时作者引用他们未曾研读的文献;我们偶尔会看到一些标题看似相关但实际内容无关的引用。可以观察到,有些科研人员通过一系列论文来提高发表数和引文数,新论文只是纠正其之前草草写就的论文中的某个错误。最终,有些文章的重要性长年未被发现。有些文章的低被引可能正是由于它的创新性太强以至于发表之初无人理解。

若要准确地对科研人员的工作进行评估,需要有合格的评估人员真正读懂论文,并撰写概要以说明作者的工作在相关领域中所处的位置。然后再由评估小组和被评估人员对这些概要进行仔细的讨论。

这样做很花时间（或许应该对外部评估人员所付出的时间进行补偿），但却是准确评估的关键所在。

最近（2007年）有一篇文章^[1]介绍了被许多大学和科研资助机构使用的科研评价方法，并建议用软件来支持这些方法。但是这样的技术支持只会使事情变得更糟，因为自动计数系统只会使上述对策变得更加难以识破。

文献[1]引发的关于计数的根本问题是，在多篇作者论文中如何分配各位作者的贡献。解决这一问题的困难之处在于，目前有多种作者排序规则并存，其中包括：

- 无论是否有贡献，团队领导列第一位；
- 无论是否有贡献，团队领导列最后一位；
- 根据贡献大小排序，贡献最大者排第一；
- 根据参与撰写的先后排序，写初稿者排第一；
- 根据作者姓名的字母排序。

为恰当地评价作者在论文中的贡献，需要询问他们（并且相信他们的回答），或将该论文与他们之前的论文进行比较。偶尔会有某篇论文的贡献巨大，以至于几位作者都值得称道。对于这个问题，没有什么机械化的解决方案可信任。文献[1]建议应将注意力集中在一组因覆盖面广而著称的“顶级”期刊。但是，实质性进展和重要贡献往往是更多发表于专业期刊和会议上的。甚至“二流”期刊发表的论文，有时也能够触发重要的研究方向，或者提供可能获得重大成果的数据。

只有当专家们仔细阅读每篇论文，才能确定一个作者的工作在其研究领域的贡献度如何。这一点对于计算机科学领域来说尤其正确。在这个领域中，新术语频繁出现，但其实就是用新名词表示相似概念。有的论文，题目看上去像是新主意，理念却只是旧点子。仅靠数论文是发现不了这些情况的。

可悲的是，当前的评价系统是自我延续的。那些被这个系统给出高度评价的人经常被要求互相评价并且评价他人。他们的地位来自这一制度，不太可能有改变它的动力。管理人员经常表现得好像只管数数，很可能是因为他们自己就是被这样评价的。

那些希望看到计算机科学进步的人，那些希望

看到计算机科学回报社会的人，都应该反对计分制评价。当你收到关于某科研人员的推荐信时，如果其中只列举发表的论文清单而没有对其学术贡献进行实质性评价，别理它！这样的推荐信只陈述了任何人都可以看到的內容。当你在招聘、职称评审或评奖委员会任职时，请认真阅读候选人的论文并认真评价其内容，并坚持要求其他成员也这么做。■

作者后记：当年我写这篇文章，主要是针对以英文为母语的计算机界人士被评价的方式。文章发表后，我很快收到其他领域科研人员的评论。他们都觉得，我的观点也适合于他们的领域。另外，以其他语言为母语的同行们和我也有合作。我很快意识到，我在文章中描述的问题是普遍存在的。我很高兴，讲中文的科研人员现在也能读到这篇文章。

作者：

大卫·罗格·班纳时（David Lorge Parnas）

时任爱尔兰利莫瑞克大学软件质量研究实验室主任、教授。

译者：



张健

CCF会士、公共政策委员会主任。中国科学院软件研究所研究员。主要研究方向为自动推理、约束求解、程序静态分析与检错、软件测试数据生成。
zj@ios.ac.cn



朱雪阳

CCF高级会员、公共政策委员会主任助理、嵌入式系统专委会委员、集成电路设计专业组委员。中国科学院软件研究所副研究员。主要研究方向为嵌入式系统设计、形式化方法、智能合约分析与验证等。
zxy@ios.ac.cn

（本文责任编辑：胡春明）

参考文献

- [1] Ren, J. and Taylor, R. Automatic and versatile publications ranking for research institutions and scholars. *Communications of the ACM* 50, 6 (June 2007), 81–85.