

·科学博客·
文/吕喆

学术论文真正的价值在哪里？

最近,《浙江大学学报》英文版编辑张月红老师在 *Nature* 上发表的一篇文章关于该学报稿件中有 31% 被 Crosscheck 工具检测出抄袭的文章成为科学网上的一大热点话题,诸人为此议论,连日来科学网的头版头条基本都是关于此事的。

如此惹火动人的话题,自然有各方人士的各种观点不断闪亮登场或者黯然谢幕,显得热闹非凡。鄙人没有精力去逐字逐句地仔细阅读有关消息,也不想做出自认为公正的判断,所以这几天只是带着学习的态度冷眼旁观。直到看了科学网的新星曾庆平老师的文章“抄袭几段话犯什么罪了!”忽然觉得其实抄袭这个问题并不是大家讨论得那么简单,也许我们需要跳出对是否“抄袭”、有多少“抄袭”这种“就事论事”的争论,从更高更远的视角来审视和思考相关的问题。

发表在学术期刊上的学术论文,毫无疑问是给业界同行们看的,其中的内容应该包含比较丰富的新信息,比如新思想、新观点、新方法、新理论、新结果、新解释等,具有供学者们和学生互相学习、切磋、借鉴和参考的作用,这应该是学术界同行们大抵都能够达成的一个基本共识。所以,学者和学生翻看杂志查找文献时,应该首选对他们自己的研究有启发和借鉴作用的文章,仔细阅读文章时也正是在重点阅读具有这样作用的部分和段落。我们有时看了一篇文章后会认为最值得记忆的就是其中的一段话或者一个图表(当然这篇文章的作者、发表杂志的名称及卷期页码也须记录,以便引用)。文章的其他部分,或许会和众多类似相关文献一起被“合并同类项”,很快就被淡忘。因此,新的、与众不同的东西才最能体现一篇学术论文的真正价值和水平,而那些趋同的或者高度相似的东西,虽然不能缺少,但价值相当有限。

然而非常不幸,现如今大家的关注焦点不在这些最有价值的部分,而在于没有价值的部分。因为最有价值的、创新的部分除非作者自己标注,否则不会被 Highlight (选中文字并涂色做出标记,编者注) 出来。而没有多少价值的部分,却很可能被 Crosscheck 之类的软件给 Highlight 出来,一旦 Highlight 了就会吸引眼球,也就会很快成为热议的焦点。



本文作者 吕喆, 哈尔滨工业大学物理系教授。图片为本文作者。

博客链接 <http://www.sciencenet.cn/u/boxcar/>
栏目主持人 王飞跃, 中国科学院自动化研究所研究员, 电子信箱: feiyue@ieee.org。

那么为什么这些本来无甚价值的地方反而会被 Crosscheck 给 Highlight 出来呢? 原因在于那些部分的语言描述, 很可能由于或抄或仿, 甚至由于“英雄所见略同”等原因而用了相同的词语句式去表述, 从而被认为是疑似抄袭甚至被认定为“抄袭”。如果一篇文章有几段被认定为“抄袭”, 那这篇文章从此背上这口“黑锅”, 不管这几段出现在哪一部分、在何种情况下出现, 都是好说不好听的, 绝对不是件光彩的事情, 甚至让文章中最闪光的那部分思想也蒙上又黑又厚的“锅底灰”, 让本该因为有所创新而容光焕发的作者也变得“灰头土脸”。这种现象概括起来就是, 论文主要部分的创新内容可能“成事不足”不够创新, 而论文次要部分(也不得不写的部分)的重复叙述却一定“败事有余”。此时学术论文体现出一种新的“价值”——检验一个人的学术诚信。

造成这种价值倒挂的吊诡现象的原因其实也很简单——版权。因为作者向出版商要求必须尊重其版权, 那么以往的学术论文作者写出的每一句话, 列出的每一个图表都要被后来者尊重, 在后续相关文章中都应该被合理、恰当地引用而不能草率滥用。这种要求显然是合法的, 但却未必足够合理。正常情况下, 科学论文的实验数据是新的, 图表也是新的, 甚至对实验现象分析解释的观点和论述方法都是全新的, 但有可能在引言部分和实验描述部分出现一

些和已有文献相同的表述, 这就会被软件认定为抄袭。如果后续跟进了人工判断并督促整改还好, 如果就此认定为抄袭, 作者可能就比窦娥还冤了。

其实有时作者很无奈, 一个领域的研究背景情况和进展介绍、实验描述, 在不同人的笔下也不会有太多种表达形式。前面人用过了的, 后来者要是不想重复, 就得搜肠刮肚去找新的描述, 难度很大。如果不幸这个领域很热门, 一年有 N 多篇文章发表, 后来者必然为此痛苦异常。长此下去, 如果学者们造不出新的表述语句, 或许只能用很简短的几句话表明自己最新的观点和看法, 那么结果会让学术论文变得像现如今流行“围脖”(微博)一样微薄了。

如此困局, 出路在何处? 希望读者们在思考之后给出良策。

~~~~~ 博友热议 ~~~~~

- * 我觉得, 需要将抄袭和引用区别对待。原封不动地拷贝别人的论文将其据为己有肯定是一种抄袭行为, 把别人的思想换一种说法表达出来也是一种抄袭行为, 而且是一种更恶劣的抄袭行为。前者容易检测, 而后者就不那么容易鉴别了。而正确引用则可以避免抄袭的嫌疑, 但过度引用则可能侵害版权。目前, Crosscheck 之类的软件似乎还做不到对上述情况进行正确区分。
- * 即使观点是新的, 但观点的论述可能会沿用以往的表述方式, 或者用同一个案例, 虽然对案例的解释会有区别。此类软件总不如人的慧眼, 眼下或许只能如此。
- * 对是否“抄袭”的标准讨论确实是挺重要的。上次美国杜克大学因为学生做作业“抄袭”, 一下子开除了 8 个中国留学生(另有十多个则获得停学或重修处分)。现在细想的话可能就是其他国家对于“抄袭”的标准同中国的标准不一样, 结果开除受处分的绝大多数都是中国留学生。
- * 鉴定是否抄袭, 在不同的段落, 有不同加权。综述前人工作部分的加权小一些。鉴定重点应放在后面的创新部分。

(责任编辑 李娜)