

# 大数据时代精细加工提高科技论文显示度

吕小红 杨开英 张蕾

《北京工业大学学报》编辑部,100124,北京

**摘要** 为了探索大数据时代科技期刊提高科技论文显示度的有效途径,针对检索系统和数据库的新发展,探讨了在编辑加工过程中,从题名、关键词、摘要、作者信息等方面入手,加工提炼,完善细节,提高科技论文显示度的方式方法。

**关键词** 科技论文;显示度;编辑加工;可控词;热词

**Improving the visibility of scientific papers in the era of large data**/LYU Xiaohong, YANG Kaiying, ZHANG Lei

**Abstract** To explore the effective ways of improving the visibility of scientific papers in the era of large data, we reviewed the new development of index system and database, and discussed the way to refine details and improve the exposure degree of scientific papers in the process of editing and processing from titles, keywords, abstracts, author information, etc.

**Keywords** scientific and technological papers; display degree; editing and processing; controlled terms; hot words

**Authors' address** Editorial Board of Journal of Beijing University of Technology, 100124, Beijing, China

**DOI:**10.16811/j.cnki.1001-4314.2018.04.010

中国数据总量正在以年均50%的速度增长,预计到2020年将占全球的21%,中国正在成为真正的数据资源大国<sup>[1]</sup>。截至2017年11月28日,在中国知网的网站(<http://www.cnki.net>)上,通过主题检索有关“脱氮”的文献,可找到2万511条结果;在百度学术(<http://xueshu.baidu.com>)上“百度一下”“脱氮”,发现共有约10.4万条相关结果。面对如此大量的文献信息,如何让读者检索到自己期刊上的论文是广大科技期刊编辑一直在思索的问题<sup>[2-5]</sup>。只有先被读者看到,才能谈得上被下载、被阅读、被引用,从而提高期刊的影响力,因此提高论文的显示度至关重要。由于技术层面的内容已相对成熟,且不能完全由编辑所掌控,而大数据时代,数据库和检索系统不断发展,技术日新月异,对编辑加工不断提出新要求;因此本文仅从编辑加工的层面,中文数据库主要以中国知网和百度、英文数据库主要以EI为例,分析如何通过精细编辑加工,使科技论文更专业、准确,达到提高科技论文显示度的目的。

## 1 题名

### 1.1 反映专业分类

题名是对科技论文内容的高度

概括,应让读者一目了然其论述的主要内容,明确其专业分类,突出其创新性成果。专业分类清楚,有利于检索特别是高级检索,即Expert Search。如《生物膜同步硝化反硝化脱氮过程中N<sub>2</sub>O的产生量及机理分析》<sup>[6]</sup>一文,知网的专业分类为:环境科学、安全科学—环境科学基础理论—环境化学,其英文题名为:The N<sub>2</sub>O Production and Mechanism Analysis of Simultaneous Nitrification and Denitrification in Biofilm Reactor, EI定义的“Main heading”为:“Denitrification”,专业分类较为明确;《潜艇用大容量蓄电池充放电过程建模与仿真》<sup>[7]</sup>一文,明确指出潜艇用蓄电池,知网的专业分类为:工业技术—电工技术—独立电源技术(直接发电)—蓄电池,既阐明所论述内容为蓄电池,又突出了特色是潜艇用蓄电池,有助于精确检索。

**1.2 尽可能使用专业热词** 题名是科技论文重要的组成部分,读者检索论文时,题名是他最常检索的内容之一。题名既要概括论文内容,简明准确,还要尽可能使用专业热词,以提高被检索到的机会。例如,《北京工业大学学报》2015年第1期刊载的《深度学习研究综述》<sup>[8]</sup>一文,因为题名中含有“深度学习”专业热词,是学科的热点,自出版以来至2017年11月16日,下载1万4842次,被引106次,居2008年以来该刊高被引论文的榜首。人工智能、机器学习、大数据、3D打印等都是相关领域的研究热词,题名中含有这些词汇容易被检索到。在百度学术等搜索引擎上进行检索时,检索结果中出现在前面的是标题中含有检索词的文献,所以题名中的专业用词需要慎重推敲。如搜索“脱氮”,可发现百度学术上的“相关热搜词”有“生物处理”“硝化”“反硝化”等,题名中含有这些热搜词,能够增加论文被检索到的机会。

## 2 关键词

**2.1 尽量使用叙词** 关键词用于表达论文的主题内容,是那些出现在论文题名、摘要、正文中的对表达论文主题具有实质意义的、重要的、关键性的词语<sup>[9]</sup>。GB/T 7713—1987《科学技术报告、学位论文和学术论文的编写格式》规定:“每篇报告、论文选取3~8个词作为关键词”,“如有可能,尽量用《汉语主题词表》等词表提供的规范词”。关键词使用叙词有利于提高检

\* 中国高校科技期刊研究会专项基金课题(CUJS 2017-014);2017年北京工业大学人文社科基金资助项目

索效果。关键词由作者给出,作者作为专业人士,更了解论文的内容与关键技术,所以用关键词检索一般会更加准确。可以要求作者尽量多给出几个关键词,另外最好有从论文层次标题、正文中选出的反映论文主题的专业词汇。图书馆的标引人员未必精通专业,他根据关键词和中图分类号确定主题,关键词使用叙词有助于标引准确。中国知网对关键词的同义词和上下位词进行了规约,从而提高了关键词的标引精度。有些期刊很重视论文词汇的规范,如《西北工业大学学报》发文称其发表的某论文中有13个2006年以后出现的EI关键词,可见其对关键词标注的重视<sup>[10]</sup>。

**2.2 反映技术领域** 关键词不应过窄、过专,也应尽量反映研究的技术领域,这样有利于提高显示度。ASME(美国机械工程师学会数据库)网站上有“Topic Collections”,“Topic Collections offer a quick and easy way to track up-to-the-minute content published by ASME on the topics that interest you”<sup>[11]</sup>,包括 Aerospace Industry、Applied Mechanics、Automotive Systems、Biomechanical Engineering 等,可以按照这些领域进行专业领域的检索。如《斜入射地震作用下地铁车站结构抗震性能分析》<sup>[12]</sup>一文,给出的关键词首先为“地下结构”,指出了论文所属技术领域;《有机材料热老化损伤非线性超声检测试验研究》<sup>[13]</sup>一文给出的关键词为:非线性超声、热老化、二次谐波、非线性损伤因子,可补充“无损检测”作为技术领域,从而增加检索到的概率。

**2.3 使用简练、明确、高频词汇** 关键词应含义明确、简练,不必用形容词来修饰,如可使用“生活污水(domestic wastewater)”作为关键词,而不必使用“实际生活污水(real domestic wastewater)”作为关键词;像“研究”“设计”这类词过于宽泛,不适合作为关键词;关键词不能仅仅从标题中摘录几个名词,如《基于构型选优决策的冗余度机器人避障研究》<sup>[14]</sup>一文的关键词为:构型选优决策、冗余度机器人、避障,全部摘自标题,不能比标题增加任何检索项,不利于提高显示度。EI能够识别缩写词COD为“Chemical oxygen demand”,并将其列为可控词(Controlled terms)。有些数据库如EI有自己的关键词,如Thesaurus,由EI数据库根据标题、摘要、关键词、参考文献等自己给出,不采用作者给出的关键词,作者给出的关键词规范则会被收入,否则就不会被采纳。关键词也应使用专业热词。在《中国科技期刊引证报告》(2017年版)<sup>[15]</sup>中有“红点指标”这一项,指论文中的关键词与所在学科排名前1%的高频关键词重合的论文所占的比例,借此反映论文与学科研究热点的重合度。专业热词多,红点

指标高,反映研究热点的论文多,显示度会高,被引用的概率也会高。

### 3 摘要

**3.1 重点突出** 摘要用来提供论文的内容梗概,简明但确切地记述论文的重要内容,具有独立性和自明性,并拥有与文献同等量的主要信息。一些索引数据库,只收录标题、摘要、关键词等二次文献,读者不阅读原文,只阅读摘要应能获得基本的信息。一般学术论文的摘要包括研究目的、方法、结果和结论4个部分,摘要中应避免进行背景、常识性内容的介绍,且不重复标题内容、不对论文内容进行评论。应重点介绍作者的工作内容、创新性成果,特别是独到的主要结论。如《基于构型选优决策的冗余度机器人避障研究》<sup>[14]</sup>一文的摘要为:“为了使冗余度机器人在障碍物环境下更好地完成末端操作任务,首先提出了一种新的基于构型选优决策的避障策略,它在保证避障和完成末端任务的同时使机器人的综合性能指标更优,然后利用3R-2P机器人的仿真及实验研究证明了所提方法的可行性。”该摘要虽然符合几项基本要素,但是过于平淡、简单,没有突出研究的创新点,缺乏特色,缺少有效检索项,不易引起读者的阅读兴趣,无法增加显示度。

**3.2 内容丰富** 中国知网等中文数据库、EI等英文数据库已实现了对二次文献的全文检索,所以题名、摘要、参考文献等的信息越丰富,有效检索项越多,检索到的概率越大,其中尤其以摘要撰写提高的余地最大。摘要里出现的专业词汇同样会被作为检索词、EI Thesaurus(工程索引叙词表)搜索到,所以摘要中的内容应尽量丰富,加大有效信息量,在篇幅允许的情况下,尽量写成长摘要,这样被检索到的概率就会增大。

EI Thesaurus包括可控词(Controlled terms)和非可控词(Uncontrolled terms),出自题名、摘要、参考文献,可见摘要的信息量越大、参考文献越多,可抽取的Thesaurus越多,论文被检索到的概率越大。在全文检索的情况下,增加有效信息量对提高显示度显得更加重要。如《生物膜同步硝化反硝化脱氮过程中 $N_2O$ 的产生量及机理分析》<sup>[6]</sup>一文中的“Controlled terms”(括号里是出现的位置)有:Batch reactors(摘要)—Biofilms(题名)—Bioreactors(在题名和关键词中有Bioreactor)—Carbon fibers(摘要)—Chemical oxygen demand(摘要中有缩写词COD)—Dissolved oxygen(摘要)—Nitrification(参考文献、题名、摘要)—Nitrogen oxides(参考文献)—Oxidation(摘要)—Wastewater(参考文献、关键词)。可见,摘要是EI Thesaurus最主要的来源,掌握好用词规律可有效提高检出率。

## 4 作者信息

作者也是数据库的检索项。中国知网《中国学术期刊(网络版)》可以进行作者、单位的检索;EI的Browse Indexes包括Author、Author affiliation。科技论文不应只考虑版式的美观而对作者人数和单位数进行限制。多作者的论文由于科研工作的延续性,可能更多地被检索,在各作者随后发表论文时都可能引用之前发表的论文,因此也可能具有更高的被引频次、更好的显示度。另外,一个作者有多个不同单位时也应标注全面,这样有助于检索的准确性和通过机构检索找到,增加检索的显示效果。

中国知网中有相关作者文献、同行关注文献、相关机构文献等。如《生物膜同步硝化反硝化脱氮过程中 $N_2O$ 的产生量及机理分析》<sup>[6]</sup>一文的作者来自北京工业大学北京市水质科学与水环境恢复工程重点实验室,知网中给出的相关机构有哈尔滨工业大学市政环境工程学院、沈阳建筑大学市政与环境工程学院、南昌大学环境科学与工程学院等,单位标注准确,就能吸引相关机构同行的更多关注。百度学术中可以进行学者的认证,可以查到其研究领域、发文情况,主要合作者、合作机构信息准确丰富,检索到的概率就更大。编辑在编辑加工时还应注意作者单位是否准确,常有作者将单位写成简称,影响检索结果,如将“北京工业大学环境与能源工程学院”写成“北京工业大学环能学院”等;北京市、教育部等的重点实验室都有固定的中英文名称,应保证与官方名称一致。

EI中像“王淑莹”的英文有“Wang Shuying”和“Wang Shu ying”,检索结果是不一样的,为提高检索效果,避免漏检,期刊应推荐作者注册开放研究人员和贡献者唯一标识符(open research and contributor ID, ORCID),并在论文中进行标注。

## 5 结束语

1)牢记提高显示度这根弦。俗话说“细节决定成败”。编辑加工中的细节对提高论文显示度起着不可忽视的作用。数字化时代,我们所面临的是海量的数据,作为科技期刊的编辑,除了应遵守编校规范、提高编校质量外,应从提高科技论文显示度着眼,在编辑加工中突出能提高显示度的细节。如多用专业热词,当专业词汇的英文有几种不同表达时应尽量采用专业上普遍认可的词汇;重视论文摘要的撰写,以及论文对参考文献的合理引用等。

2)先进技术是支撑。俗话说“创新驱动发展”。作为科技期刊人,必须顺应时代的进步,运用互联网思

维,更新出版理念,利用现代化的技术手段,采用先进的出版模式,如加强网站建设与论文推送,开展开放获取、优先数字出版、数字对象标识符标注,参与期刊群建设等,以便更有效提高科技论文的显示度。这些工作不是编辑所能决定的,但是编辑在工作中应积极建言献策、参与其中,促进期刊发展。

## 6 参考文献

- [1] 李景. 预计到2020年我国信息数据总量将占全球21% [N/OL]. 经济日报, 2016-08-29 [2017-12-02]. [http://www.ce.cn/xwzx/gnsz/gdxw/201608/29/t20160829\\_15312831.shtml](http://www.ce.cn/xwzx/gnsz/gdxw/201608/29/t20160829_15312831.shtml)
- [2] 杜秀杰, 赵大良, 葛赵青, 等. 数字出版时代如何提高科技论文的显示度[J]. 中国科技期刊研究, 2012, 23(1): 126
- [3] 付雅静, 钱俊龙. 数字出版时代提高科技期刊显示度途径的探讨[J]. 中国科技期刊研究, 2014, 25(10): 1262
- [4] 黄崇亚. 网络传播环境下如何提升科技论文的显示度 [EB/OL]. (2015-10-09)[2016-08-15]. <http://www.cujis.com/detail.asp?id=2430>
- [5] 王维朗, 吕赛英, 游滨, 等. 提升科技期刊国际显示度的途径与策略[J]. 中国科技期刊研究, 2011, 22(5): 743
- [6] 王淑莹, 张静蓉, 尚会来, 等. 生物膜同步硝化反硝化脱氮过程中 $N_2O$ 的产生量及机理分析[J]. 北京工业大学学报, 2011, 37(9): 1400
- [7] 杨昆, 欧阳光耀, 赵建华. 潜艇用大容量蓄电池充放电过程建模与仿真[J]. 北京工业大学学报, 2011, 37(9): 1325
- [8] 尹宝才, 王文通, 王立春. 深度学习研究综述[J]. 北京工业大学学报, 2015, 41(1): 48
- [9] 陈浩元. 科技书刊标准化18讲[M]. 北京: 北京师范大学出版社, 1998: 64
- [10] 胡沛泉. 《西北工业大学学报》第33卷第5期首篇朱怡安的论文英摘Key words有13个2006年以后出现的主题词[J]. 西北工业大学学报, 2015, 33(6): 1026
- [11] ASME. ASME topic collections [EB/OL]. [2016-10-03]. <http://asmedigitalcollection.asme.org/libproxy.bjut.edu.cn/collections.aspx>
- [12] 路德春, 李云, 马超, 等. 斜入射地震作用下地铁车站结构抗震性能分析[J]. 北京工业大学学报, 2016, 42(1): 87
- [13] 焦敬品, 李亮, 何存富, 等. 有机材料热老化损伤非线性超声检测试验研究[J]. 北京工业大学学报, 2016, 42(1): 24
- [14] 赵京, 刘先灿, 刘宇. 基于构型选优决策的冗余度机器人避障研究[J]. 北京工业大学学报, 2011, 37(9): 1292
- [15] 中国科学技术信息研究所. 中国科技期刊引证报告(核心版): 自然科学卷[M]. 北京: 科学技术文献出版社, 2017: 9

(2018-01-07 收稿; 2018-04-07 修回)