

学术论文关键词“泛义”现象观察与分析

戴 旻

(安徽大学 管理学院,合肥 230039)

摘 要: 关键词是学术论文的主要组成部分之一。关键词标引的规范,是影响学术论文检索效率的重要因素。但是,词义模糊宽泛,专指性差的泛义词被频繁用于关键词标引的“泛义”现象,却成为当前一个较为严重的问题。本文在调查这一现象存在状况的基础上,深入分析其主要表现形式和不良影响,进而探寻出切实可行的解决对策。

关键词: 学术论文; 关键词; 泛义词

DOI: 10.13783/j.cnki.cn41-1275/g4.2015.04.021

中图分类号: G254.2 **文献标识码:** A **文章编号:** 1008-3715(2015)04-0096-04

学术论文是研究人员在进行理论探索、实验观测后,依据其研究成果和创新见解撰写出的书面文件。学术论文是创新和科学的,同时也应是规范的,这是影响学术论文交流与传播的重要因素。早在1987年5月,国家标准局就颁发了《科学技术报告、学位论文和学术论文的编写格式(GB 7713-1987)》(以下简称“《编写格式》”),对学术论文的主要结构及撰写格式做出了统一的规定。

在学术论文主要组成部分中,关键词上承题名与摘要,下启正文,可以说是学术论文的灵魂体现。它以最为精炼和准确的词或词组来描述论文的主旨与重点,从侧面体现着作者的研究思路、采取的研究方法,以及涉及的研究领域。在相关标准的制约和规范下,标引关键词已经成为学术论文撰写过程中必不可少的步骤,但是,关键词抽取不当,错标、漏标、过度标引等情况的发生,在一定程度上影响了关键词的标引质量。这些问题中,尤以“泛义”现象最为突出,本文将以此一现象作为研究对象,在调查其存在状况的基础上,深入解析和透视其主要表现形式和不良影响,进而探寻出切实可行的解决对策。

一、学术论文关键词“泛义”现象调查

“泛义”现象是指学术论文关键词中频繁出现

诸如发展、研究、对策、问题、现状等词义模糊而宽泛,专指性差,检索价值低的泛义词的现象。这是与我国关键词标引精确性与专指性基本原则相背离的,将会降低论文检索的效率,影响文献计量的科学性。

为了进一步调查和了解“泛义”现象的存在状况,本文以中国期刊全文数据库中收录的1994年至2015年7月17日15时40分的学术论文作为调查对象,将调查范围限定在新闻与传媒、出版、图书情报与数字图书馆、档案及博物馆四个学科类目中,以“中文关键词”作为检索入口,首先统计出每个学科类目中出现频次最多的前40个关键词,统计结果如表1所示。

依据表1,笔者进一步对四个学科类目的前40个关键词进行词性与词义上的分析,找出其中的泛义关键词,统计其数量,进而计算其在40个关键词中所占比率,结果如表2所示。

由表2可知,新闻与传媒、出版、图书情报与数字图书馆和档案及博物馆四个学科类目的前40个关键词中,均有较多数量的泛义词存在,其中出版学科的泛义关键词就高达22个,约占总数的55%;档案及博物馆学科的泛义关键词也达到总数的50%。

收稿日期:2015-07-20

基金项目:国家社科青年基金项目“我国非物质文化遗产建档标准体系研究”(12CTQ038)

作者简介:戴旻(1981—),女,安徽桐城人,管理学博士,安徽大学管理学院副教授,硕士生导师,主要研究领域为档案文献遗产保护、档案信息资源建设。

表1 1994—2015年四个学科类目出现频次最多的前40个关键词

学科	论文数(篇)	关键词(括号中数字为出现频次)
新闻与传媒	621149	新媒体(6174) 微博(4154) 传播(2820) 创新(2620) 新闻(2617) 网络(2270) 对策(2230) 电视新闻(2145) 媒体(2087) 电视(1981) 受众(1756) 发展(1738) 媒介融合(1728) 主持人(1691) 网络舆情(1595) 大众传媒(1589) 新闻报道(1535) 传统媒体(1512) 影响(1436) 舆论监督(1392) 策略(1385) 民生新闻(1369) 媒介素养(1363) 电视节目(1357) 问题(1333) 大学生(1329) 互联网(1304) 媒介(1296) 网络舆论(1175) 网络媒体(1172) 舆论引导(1170) 网络传播(1116) 广播(1094) 报纸(1064) 现状(1040) 新闻传播(1034) 广播电视(1014) 突发事件(1013) 信息(1003) 传播学(991)
出版	425430	科技期刊(7165) 编辑(4945) 高校学报(2875) 学术期刊(2636) 期刊(1904) 学报(1845) 数字出版(1588) 出版(1416) 著作权(1145) 对策(1128) 发展(978) 创新(963) 质量(928) 科技论文(776) 问题(684) 编辑工作(654) 素质(643) 参考文献(626) 数字化(618) 编辑素质(612) 出版业(611) 医学期刊(597) 特色(596) 网络(552) 作者(520) 规范化(519) 现状(517) 出版社(516) 图书(515) 改革(495) 选题策划(494) 学报编辑(470) 品牌(462) 定位(455) 审稿(453) 期刊质量(434) 网络出版(433) 学术质量(431) 国际化(427) 期刊编辑(420)
图书情报与数字图书馆	427814	图书馆(58004) 高校图书馆(38883) 数字图书馆(11928) 公共图书馆(9976) 信息服务(9952) 网络环境(7552) 读者服务(6448) 服务(5178) 图书馆员(5068) 管理(4534) 信息资源(4439) 对策(3982) 高校(3908) 图书馆管理(3593) 建设(3531) 资源共享(3459) 大学生(3268) 创新(2932) 网络(2899) 数据库(2822) 知识管理(2765) 图书馆服务(2553) 学科馆员(2452) 电子阅览室(2433) 高职院校(2342) 读者(2334) 数字化(2326) 图书馆学(2261) 知识服务(2241) 服务模式(2149) 知识经济(2087) 发展(2079) 素质教育(2067) 馆员(2006) 信息检索(1950) 以人为本(1943) 统计分析(1924) 信息(1915) 素质(1915) 大学图书馆(1871)
档案及博物馆	261176	档案管理(9406) 档案(6913) 管理(5254) 博物馆(2878) 对策(2682) 信息化(2499) 问题(2150) 档案工作(1874) 人事档案(1856) 高校(1724) 电子文件(1614) 创新(1317) 电子档案(1291) 数字化(1153) 建设(1043) 利用(1041) 档案馆(1033) 措施(985) 档案信息(878) 科技档案(826) 城建档案(788) 现状(759) 思考(739) 服务(728) 信息化建设(708) 企业(705) 开发利用(675) 档案信息化(673) 企业档案(673) 发展(660) 医院(659) 作用(655) 信息资源(653) 高校档案(652) 数字档案馆(634) 档案利用(608) 现代化(574) 保护(565) 归档(530) 事业单位(506)

表2 1994—2015年四个学科类目典型泛义关键词

学科	典型泛义关键词(括号中数字为出现频次)	数量	所占比重
新闻与传媒	传播(2820) 创新(2620) 新闻(2617) 网络(2270) 对策(2230) 发展(1738) 影响(1436) 策略(1385) 问题(1333) 互联网(1304) 媒介(1296) 广播(1094) 报纸(1064) 现状(1040) 信息(1003)	15	37.5%
出版	编辑(4945) 期刊(1904) 学报(1845) 出版(1416) 对策(1128) 发展(978) 创新(963) 质量(928) 问题(684) 素质(643) 数字化(618) 特色(596) 网络(552) 作者(520) 规范化(519) 现状(517) 图书(515) 改革(495) 品牌(462) 定位(455) 审稿(453) 国际化(427)	22	55.0%
图书情报与数字图书馆	图书馆(58004) 网络环境(7552) 服务(5178) 管理(4534) 信息资源(4439) 对策(3982) 高校(3908) 建设(3531) 创新(2932) 网络(2899) 数据库(2822) 读者(2334) 数字化(2326) 发展(2079) 馆员(2006) 统计分析(1924) 信息(1915) 素质(1915)	18	45.0%
档案及博物馆	档案(6913) 管理(5254) 对策(2682) 信息化(2499) 问题(2150) 高校(1724) 数字化(1153) 建设(1043) 利用(1041) 措施(985) 现状(759) 思考(739) 服务(728) 企业(705) 发展(660) 医院(659) 作用(655) 保护(565) 归档(530) 事业单位(506)	20	50.0%

除了数量较多外,这些泛义词在本学科学术论文中出现的频次也很高,有的甚至超过本学科的专业词组。如“传播”一词,在新闻与传媒学科内出现频次高达2820次,在该学科高频关键词中位列第三。有的泛义词,如“发展”“对策”“创新”等,因其词义宽泛,可以替代其他具体词语,故而突破学科界限,成为四个学科类目共有的关键词。可见,“泛义”现象已经成为学术论文关键词标引中一种较为

显著的现象。

二、学术论文关键词“泛义”现象解析

在量化统计的基础上,笔者围绕四个学科类目中出现的泛义关键词,以及出现这些关键词的学术论文做了进一步的检索和分析,进而定性地总结出“泛义”现象的主要表现形式。

第一,《编制格式》规定,学术论文关键词的数量一般应保持在3—8个之间。但是部分学术论文

标引出的关键词数量较少,未达到国家规定的要求。为此,部分学术论文不得不抽取泛义词作为关键词,以扩充关键词标引的数量^[1]。如《高校图书馆统计指标体系的构架与规范化探讨》一文,其标引的关键词为“高校图书馆;统计指标体系;网络化”,其中的“网络化”就是属于泛义词。

第二,《编制格式》指出“关键词是为了文献标引工作从报告、论文中选取出来用以表示全文主题内容信息款目的单词或术语。”^[2]因此,学术论文的题目、摘要以及正文共同构成了关键词的来源。但在具体标引过程中,切分题目直接获取关键词的情况十分普遍。如《档案资料的归纳与利用及档案工作创新的意义与途径》一文,其关键词就为“归纳;利用;创新”,很显然,该文的关键词仅仅罗列出题目中三个主要泛义动词——“归纳”“利用”和“创新”,而遗漏了该文的研究主体“档案工作”和“档案资料”,无法明确表达该文主题意义,这是不科学的。

第三,如前所述,学术论文关键词标引时,精确性与专指性是其必须遵守的基本原则。但是,现行的部分学术论文中,随意拆分、拼接词或词组组成关键词的情况非常严重^[3],拆分和拼接后的词组,有的已经失去了本义,有的与原文主题不相符,在造成概念误解的同时,也产生了更多的泛义关键词。如《网络新闻标题的语言特色》一文,其关键词为“网络新闻;标题;语言;特色”,“语言特色”原本是一个词,却被拆分成两个独立的泛义单词,也没有体现出该文的主题概念。

第四,关键词有着不同的词性和词义,它们的关键程度,以及与论文主题的相关度都不相同。2009年颁布的《文献叙词标引规则(GB/T 3860-2009)》提出了先主体因素,次通用因素,再空间因素,最后才是文献类型因素的排列顺序^[4],以指导关键词标引的实践。但是,部分学术论文并未依据这一顺序排列,导致“轻重失当”局面的出现。如《科技期刊特色感悟》一文,其关键词为“特色;科技期刊;生态位;感悟”,其中泛义词“特色”就取代主体词“科技期刊”。

三、学术论文关键词“泛义”现象透视

学术论文关键词“泛义”现象的出现,反映出我国当前学术论文关键词标引质量的低下,势必会带来一些不良的影响。一方面,一些通用的泛义词,如“发展”“问题”“创新”等,它们不仅以较高的频次出现,而且还出现在不同学科、不同领域之中。以这些泛义词进行检索,无法达到依据一个关键词提示

某一论文主题内容的专指性要求,漏检率和错检率将会增加。以泛义关键词“发展”为例,仍将检索范围限定在上述四个学科类目之下,以“发展”进行关键词检索,得到相关度最高的前5篇论文,分别是《新中国体育学术(科技)期刊发展研究》《媒体语言发展刍议》《公共文化服务体系中的图书馆定位和发展分析》《网络问政及其发展态势探微》《我国开放存取发展现状和建设策略研究》,文章分属于不同的学科领域,涉及的研究方向也各不相同,无法满足检索的需求。

另一方面,抽取的关键词过于空泛,常常会削弱对论文主题的确切表达,使这些论文的漏检率增加,被引频次降低。以《博客接受模型:影响用户接受和更新博客的实证研究》一文为例,该文于2009年4月发表于图书情报文献学CSSCI检索刊物《情报理论与实践》杂志上,该文以信息技术用户研究领域中最具影响力的理论——技术接受理论为理论依据,对用户接受和更新博客的意愿进行实证研究,是一篇颇有价值的学术论文。该文刊载至今已有7年,被引频次只有9次,笔者认为,抽取的关键词不“关键”可能是造成这一问题的主要原因。该文的关键词是“博客;理论;模型;实证研究”,除了“博客”一词外,其他三个词都是泛指性比较强,没有实质意义的词语。而发表于2009年7月图书情报文献学CSSCI检索刊物《情报杂志》上,与前文采用相同研究方法的《移动商务消费者初始信任影响因素的实证研究》一文,其关键词就分别为“移动商务;初始信任;技术接受”,“移动商务”是研究对象,“初始信任”是研究内容,“技术接受”则是理论依据和研究方法,该文标引的关键词准确地反映出该文研究最主要、最核心的内容,同时,也保证了其检全率和检准率,因此其被引频次已达到31次,远远高于前文。被引频次是关系到论文、著者,以及载文期刊学术影响力的重要因素,因此,必须要对这一问题给予高度的重视。

四、学术论文关键词“泛义”现象对策探析

学术论文关键词“泛义”现象的出现,从某种意义上反映出我国当前学术论文关键词标引的随意与不规范,其中固然有扩大论文检索入口的客观要求,但更多的原因,还是标引人员对关键词重要程度认识不深,自身知识水平和操作能力有限,以及对标引规则的不了解。针对关键词标引的不规范,众多的研究人员都提出要强化关键词重要性的认识,提升标引人员水平。^[5]这些对于规避和建设“泛义”现象有着普遍的指导作用。但是,专门针对“泛义”现象

的解决对策,目前还没有明确提出。因此,笔者认为,在参考前文“泛义”现象表现形式的基础上,规避和减少“泛义”现象,需要从关键词标引规范化和关键词质量控制两个方面做起。

关键词标引规范化方面:国家的相关标准规定,标引的关键词,应尽量选用论文所属学科范畴之内,义项较为单一的词。因此,标引关键词时,应尽量使用《汉语主题词表》或者《中国分类主题词表》中的规范词,不用或少用泛义词。^[6]即使非要使用泛义词,也可以进行合理的组配,选用通用因素与主体因素相结合的自由词,以提高关键词的专指度。关键词来源于题目、摘要和正文,在抽取关键词时,应先从题目中挑选,当题目无法涵盖论文全部主题时,需从摘要,甚至正文中抽取,避免简单切分题目,以偏盖全,也不允许停留于表面,需要重视对学术论文深层内涵的提炼。此外,还应该合理控制关键词数量,避免泛义关键词的堆砌而引起的信息冗余,依据论文主题确定标引深度,以求在主题内容与检索效率之间保持平衡。

关键词质量控制方面:从学术论文关键词抽取和标引的过程来看,著者是关键词的第一提供者,期刊编辑及审稿人是关键词的第一审核和修改者,期刊全文数据库标引人员则是关键词的最后把关人员。但从本文的调查结果来看,本调查所依托的中国期刊全文数据库,采用的就是不依据词表,由作者根据论文主题内容自主拟词的自由标引方式^[7],这在一定程度上将会增加关键词标引的随意性。除了中国期刊全文数据库,维普中文科技期刊数据库和万方数据知识服务平台也采取的是自由标引方式,对著者标引的关键词几乎不作改动,从而造成关键词标引后期质量控制的缺失。因此,必须引起国家

权威部门、期刊全文数据库开发商、期刊编辑和著者的高度重视,只有充分发挥著者、编辑、审稿人、数据库标引人员的共同控制和监督作用,才能确保和提升关键词的标引质量。

总之,“泛义”现象的存在,体现出当前我国学术论文关键词标引中存在的问题与不足,同时也指明了下一阶段关键词标引规范化工作的主要着力点。只有科学、全面、准确地抽取、拣选关键词,才能确保关键词标引的质量,提升学术论文规范化水平,帮助用户更加方便有效地利用文献,掌握文献的主题内容。

参考文献:

- [1]赵妍,侯汉清.中文期刊文献通用词标引分析[J].图书与情报,2007(1):64.
- [2]GB 7713-1987 科学技术报告、学位论文和学术论文的编写格式_百度文库[EB/OL]. [2015-07-11]. <http://wenku.baidu.com/view/b4a1da19a300a6c30c229fca.html>.
- [3]徐雨衡.学术论文摘要的写作之道[J].出版发行研究,2012(2):62.
- [4]文献叙词标引规则(GB/T3860-2009)[S]. [2015-08-12]. <http://bzwxw.com/download/06/GBT-3860-2009ZLH.rar>.
- [5]倪向阳,马永军.科技期刊的关键词标引质量亟待提高:从《编辑学报》《中国科技期刊研究》的标引现状谈起[J].编辑学报,2011(4):307.
- [6]吴立志.学术论文关键词的概念及标引方法辨析[J].现代情报,2009(6):9.
- [7]董丽,侯汉清.中文期刊文献关键词标引的分析和改进[J].情报科学,2004(11):1356.

(责任编辑 吕志远)

Observation and Analysis of Semantic Overgeneralization of Key Words in Academic Papers

DAI Yang

(School of Management, Anhui University, Hefei 230039, China)

Abstract: Key words are one of the crucial parts of academic papers. Keywords indexing standard will be an important factor to influence the efficiency of retrieval of academic papers. However, the phenomenon that words with broad and vague meaning are frequently used as keywords in academic papers has become a more serious problem at present. On the basis of investigating this phenomenon, this paper analyzes its main forms and adverse effects, and then explores the feasible solutions.

Key words: academic papers; keywords; words with universal meaning