

科技论文关键词标引应琢磨的几个问题

程智强 江洪涛

上海市农业科学院农业科技信息研究所, 201403, 上海

摘要 讨论关键词标引中几个共同关心的问题。认为:关键词选用不必拘泥文章自身,也不必计较是叙词还是自由词,但要琢磨读者检索文献时的思路;在准确反映文章主题的前提下,优先选取科技文献中使用频率高的词或词组,慎用短语;关键词排列不必死抠规则或逻辑,把关系密切的词相邻排列即可。

关键词 关键词标引;科技论文;科技期刊;文献检索

Several issues to be pondered in keyword indexing of sci-tech papers// CHENG Zhiqiang, JIANG Hongtao

Abstract Several issues of common concern in keyword indexing are discussed, from which it is inferred that the keywords do not have to be selected rigidly from the indexed sci-tech paper and also they are not necessarily argued to be descriptors or not, but ideas for readers to retrieve literature must be considered. Words frequently used in literature can be selected with priority provided they accurately reflect the paper's subject. It is also suggested that phrasal groups be cautiously used as keywords and the arrangement of keywords not necessarily stick to certain rules or logic as long as the closely related words are placed adjacently.

Keywords keyword indexing; sci-tech paper; sci-tech periodical; literature retrieval

Authors' address Agricultural Information Institute of Science and Technology, Shanghai Academy of Agricultural Sciences, 201403, Shanghai, China

自 20 世纪 90 年代以来,基于电子网络技术的计算机文献数据库建设在我国发展迅速,现在网络文献数据库(如万方数据知识服务平台、中国知网、维普网等)已成为我国科技人员检索中文科技文献的主要窗口;而关键词检索,因使用自然语言,检索灵活易学,成了读者进行文献检索的主要方法。基于这一现实,为提升科技期刊的影响力,编辑人员在努力挖掘优质稿源的同时,已越来越重视文章关键词的编辑加工。准确标引关键词,可让读者轻而易举地查到想要的文献,使自己的期刊能被广泛利用,从而提升期刊的影响力。只是现在对文章关键词的编辑加工尚难形成统一标准,因认识上的分歧,不同期刊各行其是,效果可能大相径庭。就此提出以下几个问题来讨论,供有心者思考。

1 关键词的出处与标引目标

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^[1]。毋庸置疑,关键词标引在于表达文章主题以满足情报检索需要,其标引目标自然是让文献检索者尽可能地达到高检准率和高检全率;因此,我们以为关键词出自哪里并不重要,关键是要表达文章主题并达到标引目标。也就是说,关键词的选取不必受文章自身限制,能较好地反映文章主题的名词或词组,只要权威专业词典收录,说明该词语为专业人员所认可,作者或编者就应该考虑选用。

例如《上海市城镇绿化植物规划》一文(载于《上海农业学报》2009 年第 4 期),其主题是城镇绿化及其树种规划(主要论及树种选择),虽然题名、摘要和正文均未提及“景观植物”(或“观赏植物”);但文中讨论的主要对象是景观植物(或观赏植物),且它为《英汉农业大词典》^[2]所收录,即 landscape plant——景观植物(或 ornamental plant——观赏植物),所以应把它作为该文关键词标引。至于景观植物和观赏植物哪一个作为关键词更合适,得看哪一个在科技文献中使用频率更高;因为标引目标就是要让检索者获得高检全率。显然,“观赏植物”优于“景观植物”,因为前者的使用频率远高于后者。

2 叙词与自由词的选用

叙词,原称主题词,是从自然语言中优选出来并经过规范化的词,它在标引与检索文献时用于表达各种主题概念^[3-4]。叙词标引按叙词语言的规范要求,严格遵循概念组配的原则进行,检索效率高;因此,《中国学术期刊(光盘版)检索与评价数据规范》要求对论文进行叙词标引,即按 GB/T 3860 的原则和方法参照各种词表和工具书选取关键词^[5]。

现在的问题是:叙词表(主题词表)修订周期长,有时难以反映科学技术的新进展或科技术语在运用上的新变化;更为要紧的是一般文章作者和检索文献资料的成千上万读者很少接触叙词表,要求他们查阅主题词表并了解叙词标引规则几乎是不可能的。而自由词,大多来自于文章题名、摘要和正文,我们以为只要准确地反映了文章主题,并在相关文献中有较高使用频率,就可以把它作为关键词标引。

例如,有关动植物育种的文章常把“选育”作一关键词标引,没有使用叙词“选择育种”^[6]。我们以为这

是合理的,理由有三:两者含义没有实质性差别;两者英译均是 selective breeding;更为重要的是前者使用频率远高于后者,虽然后者是规范术语。

再以上述《上海市城镇绿化植物规划》一文为例,文中提及的“植物景观”,虽不是叙词但可作为该文关键词标引;因为树种规划的目的就在于呈现良好的植物景观,而且该词也被《英汉农业大词典》^[2]收录,即 vegetative landscape——植物景观,在相关科技文献中使用频率也颇高。

总之,在准确反映文章主题的条件下,不必计较是叙词还是自由词,选用在科技文献中使用频率较高的词,即可提高文章的被检索率。

3 词汇的使用频率与关键词的选取

前面一再提及,选用的关键词要反映文章主题,要选用在相关科技文献中使用频率较高的。前者为的是确保读者检索科技文献时能查准,后者则是为了尽可能方便读者查全;但在关键词选取中有时不易同时兼顾检准率和检全率。

举2个极端的例子。

1)若把整个论文题名作为关键词,一般说来,一查一个准,但有谁想到这样的关键词呢?造成文章漏检是显而易见的。

2)若把使用频率高但主题专指度低的词汇(如综述、现状、应用、对策等)作为关键词,会发生大量误检,一些风牛马不相及的文章会同时被搜索出来。

前一例的关键词自然不会被认可,而后一例的关键词应慎用而不是一概排斥。我们以为,当已有专指的通用概念词被列入关键词时,可以把这些使用频率高但专指度低的词汇也列为关键词,如在论文关键词标引个数较少情况下,它对读者检索文献是有帮助的。通过关键词组合检索,读者可较为准确地查到想要的某一方面的文献资料。

顺便指出,关键词是词或词组,短语应慎用。为了提高关键词的专指度,不少作者把关键词写成短语,如“关键词选取准则”^[7],几年过去了,用这关键词在中国知网上只检出作者自己的一篇文章;因此,短语用作关键词,就得考量它的使用频率,除非是新生事物。

4 读者检索角度与关键词标引

大家知道,农业生产主要是动植物的再生产,在这过程中或多或少会遭遇不良环境的影响,如低温灾害、高温危害、旱灾、水灾、湿害、风害、病虫害,等等。就农作物而言,反映作物对上述逆境适应能力的农艺性状就有抗寒性、抗热性、抗旱性、抗涝性、耐湿性、抗风性、

抗病性、抗虫性,等等;针对盐碱地、贫瘠土壤,则有抗盐性、抗碱性、耐瘠薄性。读者在检索有关作物文献时,更多想到的自然是作物自身的抗逆性如何,即用相关的抗性 or 耐性关键词进行检索;再者,降低逆境影响的农艺措施如何,即用相关的农业控制方法关键词进行检索,以了解相关的情报资料。

因此,在关键词标引以及编辑加工时,文章的作者或编者理应关注或迎合读者检索文献时的思路和习惯,使自己所写论文或所编期刊更多地被利用。

5 英汉(或汉英)专业词典与关键词规范

英文是国际上强势语种,大量先进的科技文章是以英文发表的。一个中文科技术语,一般总有一个英文词语与之对应。事实上,好多中文科技术语来自于英文科技术语的直译、意译或音译,因此,未列入权威英汉专业工具书中的词语,应慎用或不用。

例如,“水稻生长习性”,在《英汉农业大词典》^[2]中可找到 rice(水稻),growth habit(生长习性),却没有 rice growth habit、growth habit of rice 或 habit of rice growth。再如,“DNA 的高效提取”“二倍体和四倍体杂交”等等,除非该词语已列为叙词或在文献中使用频率颇高。

就农业领域而言,涉及动植物育种、种植或养殖、设施园艺、生态环境、园林绿化、生物工程、农业气象、经济生产、农业机械,等等,范围之广,即使资深农业科技编辑也不可能对以上各专业的科学术语都有较深的了解;因此,借助于权威英汉专业词典来规范关键词标引,无疑是不二选择。

6 关键词排列顺序

要把选用的多个关键词列出来,就有先后的问题。有人认为,关键词应按其“关键性”递减的顺序排列。事实上,几个关键词有时很难说清哪一个更关键。也有人强调关键词排列按逻辑关系,或按研究目的→研究类别→研究方法→研究结果的顺序排列^[8]。可是,对综合性科技期刊而言,它的文章包括学术论文、研究报告、科研简报、专题综述等不同类型,很难套用同一个模式。

我们以为,作者和编者只要把专指度高的关键词大致列前并把关系密切的词(如属一种、整体一部分)尽可能相邻排列就行了。对检索者或读者而言,关键词只是检索的入口词,放在前与列在后检索效果没有多大差别。找到文章后,读者要了解文章的内容就直接看题名和摘要,不会再根据关键词的排列顺序去分析文章内容。