

文献主题标引的误差及其控制

刘 继 红

(图书馆)

摘 要 就主题分析和主题标引工作中常常出现的误差进行了较全面的分析,并提出了实行标引质量控制的可行性措施。

关键词 主题分析;主题标引;误差分析;控制

分类号 G254. 2

文献标引工作是文献情报检索与利用的前期工作,文献标引的准确性是保证较高的文献检索效率和利用率的关键性环节。然而,在实际工作中,由于标引人员的知识水平、思维方式和标引技能等主观因素及某些客观因素,导致标引错误时常发生,影响了文献标引的质量,降低了主题检索系统的功能。所以,本文就文献主题分析和主题标引工作中常常出现的误差进行了较全面的分析,并提出了实行标引质量控制的可行性措施。

1 主题分析误差

主题分析是标引人员对文献的主题进行概念分解和概念综合的过程。掌握文献或课题中究竟包含哪些有检索意义的主题概念,以便进行精选和取舍。选定应标引的主题概念,为下一步主题标引工作打下良好的基础。在主题分析过程中常常出现下列误差:

(1) 主题概念提炼不全。这是指从文献中提炼和分析出来的主题概念少于原文献提供的主题概念,从而把有价值的主题概念漏掉,导致漏检,影响文献的查全率。例如,“张友渔学术论著”一书,分法学、新闻学、政治学和国际政治四个部分。若只就题名分析出“张友渔”、“文集”二个主题,就把“法学”、“新闻学”、“政治学”与“国际政治”这些重要的概念漏掉了。

(2) 主题概念提炼过多。这是指从文献中提炼和分析出来的主题概念多于原文献提供的主题概念,这种现象把检索意义不大的主题概念作为标引词,容易造成在同一标引深度上重复标引,影响查准率和查全率。例如,“外语教学中培养非智力因素的研究”一文,正确的标引词是:“外语教学”、“教学改革”、“非智力因素”、“培养”、“中学”。若再加上“外语课”、

收稿日期:1994-12-03.

刘继红,女,1948年生,馆员。

“课堂教学”二个主题概念，就是过多地重复标引了。

(3) 主题概念提炼错误。这是指从文献中提炼和分析出来的主题概念与原文献不符。这种现象将引起错标，导致错检、漏检。例如，“现代建筑画选”一书，该书是反映建筑艺术方面的画册，不涉及建筑设计，正确的标引词是，“建筑艺术”、“绘画”、“中国”、“现代”、“画册”。若将“建筑艺术”改标为“建筑设计”就属于错标了。

(4) 主题概念取舍错误。这是指标引人员对提炼和分析出来的主题概念，在选定应标引的主题概念时，该省略的没省略，不该省的却省了。例如，“冯至学术论著自选集”一书，该书收集了作者39篇重要学术文章，介绍评述了世界文化名人及其作品。正确的标引词是：“冯至”、“文学理论”、“文学研究”、“世界”、“文集”。若标为：“冯至”、“作家”、“世界”、“文集”，则舍去了反映文献核心内容的主题概念，选取了不应该选的或可以省略的主题词。

(5) 主题概念提炼性不专指。主题分析的专指性是指从文献中提炼出来的每一个主题概念的外延和内涵必须与文献论述的每一个主题概念的外延和内涵相符。也就是说从文献中提炼出来的主题概念在外延上不论是过大还是过小，在内涵上不论是过深还是过浅，都是非专指性的。例如，“中国经济体制改革”一书，若提出主题“中国体制改革”或“经济改革”其外延都大于原文献外延，内涵都过浅，提炼概念时违背了专指性原则。

(6) 主题概念取舍性不专指。文献中提炼出来的复杂主题包含的主题因素也相对多。在实际标引中，由于受标引深度限制，需舍去一些次要因素，在取舍过程中如把握不准，就会影响标引的专指性。例如，“CC—DOS 操作系统技术手册”一书，标引词应该是：“汉字处理软件系统”、“磁盘”、“操作系统”、“手册”。“手册”一词视具体情况也可舍去不标，但其它词不能省略。

2 主题标引误差

主题分析后选定需要标引的主题概念，然后将这些概念一一转换为词表上规范化的主题词。如果标引工作的第一步——主题分析做得尽善尽美，但主题词转换错了，也不可能得到正确的检索标引。相对而言，在标引工作所有的环节中影响标主题词的因素最多，也可以说标主题词是标引步骤中最困难的一步，标引误差常在此产生，常产生的标引误差如下：

(1) 错标。标引人员或因缺乏专业知识或未对文献进行全面仔细的分析，对专深的技术文献，往往把握不住其核心概念，对所标引的概念不甚了解，结果导致错标。

(2) 漏标。是指在对文献主题进行分析时，对其概念提炼不全，把有检索价值的主题概念漏掉，导致漏标。这些问题常常出现在对隐性主题的标引上。例如，“高温下使用的不锈钢”这一概念，“不锈钢”是显性主题，而“耐热钢”是它的隐性主题。若稍加不慎就会把“耐热钢”这一隐性主题漏标。

(3) 过度标引。标引人员有时摸不准文献中需要标和不需要标的主题，以为选择的主题词越多越好，或在同一标引深度上反复标，或选用与文献关联不太大的主题进行标引，其结果导致过度标引。（例子与1（2）同）

(4) 过粗标引。主要是指违反了标引专指原则。标引选词应该是与原文献主题概念相对应的、单个的最专指的主题词，只要词表中有这个专指词存在，就不得用其上位或下位概念词，

也不能进行组配。如,“邓小平社会主义经济思想论”一主题,标引词是:“社会主义政治经济学”、“邓小平”、“思想评论”。若标为:“政治经济学”、“邓小平”、“思想评论”就属过粗标引了。

(5) 组配标引误差。组配是主题标引技术与主题索引技术中的基本手段,组配技术灵活是主题法最基本的特点。然而组配是最容易产生标引误差的环节,常见的组配误差如下:

1) 字面组配误差。是指单纯从字面的拼合形式出发,而不考虑概念之间的关系和语法关系的组配。例如“海燕收录机”一主题,若采用“海燕”与“收录机”组配就是单纯的字面组配。正确的组配应是:“海燕牌商品”和“收录机”组配,这样标引和检索出来的文献才不至于把有关“海燕”这种鸟类的文献与“收录机”的文献混杂在一起。

2) 二义性组配误差。二义性组配的含义是指一组组配标识,其组配结果可能同时表达几个不同意义的主题。例如,“体育学校现状”一主题,用“体育”和“学校”这一组组配标识,其组配结果可以表达和描述两个性质完全不同的主题。正确的标引应是:“体育—学校—现状”,不能标为学校—体育—现状。

3) 漏义组配误差。在组配标引时,选用的组配主题词比原文献表达的主题概念含义要少而造成的误差。例如,“离心式泥浆泵的系数设计法”一主题,标引词应是:“离心泵”、“泥浆泵”、“设计”、“数学方法”。但在标引中只选用了“泥浆泵”、“设计”、“数学方法”几个词组配,而把“离心泵”这一重要的主题词漏掉了,造成漏义组配误差。

4) 越级组配误差。越级组配是指在标引文献时,凡以粗代细,即以泛指词代替专指词;或以细代粗,即以专指词代替泛指词的组配都是越级组配。例如,“日本对华政策”一主题,若选用“对华政策”、“日本”二个词组配是正确的,若选用“对外政策”、“日本”二个词组配就属于越级组配错误了。

5) 虚假组配误差。虚假组配是指在标引时一篇文献的几组并列主题之间的虚假联系所形成的一种组配,这也是造成检索误差、影响查准率的重要原因。例如,“棉花的种植和玉米的管理”一文献,论述的是两个并列的主题,标引词是“棉花”、“玉米”、“种植”、“管理”。若不采用职能符号加以限定,而将标引词轮排组合可产生象棉花—玉米,种植—管理等一些毫无意义的虚假组配。

6) 能用交叉组配时却采用方面组配而造成的误差。依据标引规则规定,在可进行交叉组配又可进行方面组配时,要优先使用交叉组配。例如,“硬质合金不重磨刀具”一主题,若词表中不仅有主题词“硬质合金”、“不重磨刀具”,还有“硬质合金刀具”一词的话,正确的组配方案是用“硬质合金刀具”与“不重磨刀具”组配,而不用“硬质合金”与“不重磨刀具”组配,因为前者是交叉组配,后者是方面组配。

(6) 标引人员未充分利用词表,对词义判断不当所造成的误差。利用词表进行概念转换时,往往会出现一些错误,错误原因或是没有充分利用词表的各种语义参照系统,或是未认真地查阅词表,或对专业术语缺乏了解,结果导致错标。

(7) 标引不一致误差。标引人员的观点、方法和知识水平,存在差异。同一标引人员在不同的时间里对同一文献的标引也存在差别,标引人员的这些差异性,直接影响了文献标引的质量。

3 文献主题标引误差可能性控制

(1) 提高标引人员的素质。人的因素是主题标引关键性环节的重要保证。只有在保证标引人员相对稳定的前提下，全面提高其素质，包括政治素质、职业道德素质、专业知识和相关学科的知识水平，才能胜任主题标引工作。这一根本性的问题解决了，才能谈其它问题。

(2) 集中标引。集中标引是指在全国范围内指定一专门的权威性机构，担任全国文献统一标引工作。集中标引是解决文献标引一致性及其控制的最有效途径。

(3) 制定主题分析提纲。主题分析的质量和水平是影响标引准确性的重要因素。为保证主题分析的客观性、专指性和一致性，可根据不同文献机构的不同分析要求，不同的分析项目，制定主题分析提纲，供主题分析使用，以克服任意性。

(4) 制定并完善标引规则。以主题法理论、标引总则为依据，可根据实际工作需要制定本单位标引细则、标引深度。对在标引中容易出现差错的环节要严加控制，并在实践中不断地修订和完善，做好标引工作规范化。

(5) 加强对《词表》的科学管理。任何一种工具都要在使用中经常维修管理和更新改进。科学技术日新月异，新事物、新概念不断涌现，只有及时地、不断地修整词表，才能使其具有生命力和使用价值。鉴于词表庞大，不易更新，建议有关管理机构定期出版“新增主题词”小册子，及时发给使用单位，同时还应尽快出版各专业主题词表，以提高专业性较强的文献标引的准确性和一致性。

(6) 建立“固定组配结构形式档案”。在词表还未收入正式主题词的情况下，对学科专业中出现的新概念主题，可采用组配的标引方法，并建立“固定组配结构形式档案”。如，“国际经济法”一概念标为国际经济关系—经济法；“商品信息学”一概念标为经济信息学—商品学。以保证组配标引的准确性与一致性。

(7) 加强自审和互审工作。校对审核是标引工作减少错误提高标引质量不可缺少的一环。通过自审与互审工作可及时发现并纠正标引工作存在的各种错误和疏忽，提高标引工作质量。

参 考 文 献

- 1 刘湘生. 主题法的理论与标引. 北京: 书目文献出版社, 1985. 307
- 2 钱起森. 汉语主题词表标引手册. 北京: 科学文献出版社, 1985. 69
- 3 李凤满. 文献主题标引非专指性分析. 情报资料工作, 1992, (2): 21