

李洪建 顾潇华

关于检索用户培训中“主题分析”问题的思考^{*}

摘要 “主题分析”是检索工具或数据库标引人员和检索人员在检索标识上取得一致的关键。检索用户培训教材中应专设“主题分析”章节。内容应包括:主题分析的基本问题、课题的主题类型分析、主题结构分析及主题分析方法等。参考文献4。

关键词 用户培训 主题分析 检索系统

分类号 G252.7

ABSTRACT “Subject analysis” is a key for the consistency of indexing workers and search users of search tools or databases. In this paper, the authors think that “subject analysis” should be included in the textbooks for search users. Besides, details of “subject analysis” are also specified. 4 refs.

KEY WORDS User training. Subject analysis. Search system.

CLASS NUMBER G252.7

1 对主题分析在检索过程中作用的再认识

1.1 信息的标引与检索

标引和检索是一个事物的两方面。一方面标引人员将欲序化信息按一定规则编入检索系统时,必须对信息内容进行主题分析,并判断所属学科,进而转换成分类、主题语言等检索标识,将相关信息按检索标识顺序排入检索系统的正排或倒排档中(印刷型检索工具称为正文和索引)。这样,检索工具就揭示了经标引人员整理后对用户有用的信息,否则,用户所需信息将处于一种无序的原生分布状态,查找起来费时费力,达不到全、准、快、便、省的检索要求^[1]。另一方面,用户利用检索系统来满足自己的信息需求必须使用检索语言,必须对自己的课题进行主题分析,将反映课题的所属学科和主体因素及其他因素转换成相应的检索标识后,才能使用按检索语言(或称检索标识)编排的检索系统查检所需信息线索、摘要或全文数据等。

从这两个方面我们不难看出,主题分析不仅是标引而且也是检索的关键环节。主题分析正确得当,学科和主题概念就容易提炼。而主题概念提炼正确就会比较容易转换成精确、标准的检索标识,从而提高标引质量和检索效果。

在信息资源管理中,基于图书资料、档案管理的分类标引和主题标引的规范化,是在其标引过程中由相应的标引规则(分类法及使用手册、主题词表及使用手册等)来控制。用户按相应的规则,对自己的研究课题进行主题分析应与标引人员对同样内容的信息所做的主题分析结果趋于一致,这样所转换的检索标识才能一致,检索的结果才可能准

确。因此,培训检索用户的“主题分析”能力,是检索用户培训课程中相当重要的一部分,但就目前而言也是比较薄弱的环节。

1.2 对检索用户培训中“一串钥匙”与“一把钥匙”关系的新认识——主题分析是“一把钥匙”培训法

自文献检索教育开始,就有两种用户培训方法。其一,交给用户“一串钥匙”,即一种一种地介绍各种检索工具的使用方法,让用户掌握“一串钥匙”来打开知识的宝库。其二,交给用户“一把钥匙”,即只介绍检索语言,用户掌握了检索语言就可以使用任何检索工具和机读数据库,来打开知识宝库。笔者以为,“一串钥匙”法往往只能让用户对检索工具知其然不知其所以然,因此,每当遇到新的检索工具时都会产生使用上的困难,不利于从用户的使用角度对检索工具提出优化和更高的要求。而从主题分析着手,让检索用户学会用“一把钥匙”去打开知识宝库的大门,是从逆向思维解决检索语言难关的一条途径。掌握了主题分析方法,用户可以比较容易地理解任何检索工具的编排和使用方法。当然,前提是在检索用户培训中必须讲清楚“主题分析”这一检索步骤,从用户的专业知识角度讲明白检索语言。否则,问题仍将不可能解决。

2 目前我国检索用户培训中对“主题分析”的处理方式

2.1 介绍检索步骤时,“主题分析”作为第一步的方式

检索过程可分成若干步骤,在检索用户培训时必须将检

^{*} 本文是2001年河北大学教学改革课题“信息管理与信息系统专业课程教学内容与教材建设”成果之一。

索步骤分割开来,一一陈述,做到条理清晰、易于理解。检索步骤的第一步是主题分析,或称明确检索要求。一般是将主题分析分成几项内容:分析课题所需信息的所属学科;明确相关学科;分析检索范围(筛选文献信息的初步设想);明确课题所属类目及主题词(同义词、外文词、学科专业名称或代码等等);确定查找信息的文献类型;确定查找信息的年代;确定特定人物。这种分析方式看似已经清楚地介绍了主题分析的方法,但在实际运用时,用户仍感到困惑,比如:课题所属类目如何明确,明确的依据是什么?课题的主题词又应如何判断?导致对主题分析方法在信息检索中究竟有多大作用往往不甚理解。

2.2 介绍具体检索工具的使用方法时,在实例中附带分类范围和主题概念语词的分析

有些检索用户培训教材,在其检索步骤章节只是比较简单地介绍主题分析,至于如何具体做出主题分析结果,是在讲到某一检索工具时,用举例说明。如陈界老师等人主编、中国科学技术出版社出版的《医学信息检索与利用》一书就是采取的这种方法。该书涉及的每种检索工具使用方法举例中,都讲述某一课题的主题分析方法和分析结果。比如在介绍《中文科技资料目录(医药卫生)》时,其检索实例就是“骨肿瘤的放射治疗”。进行主题分析时,介绍该课题的主题词为“骨肿瘤”与副主题词为“放射治疗”^[2]。由于该课题涉及学科范围较窄,专业性又很强,按类浏览筛选量小,所以就没有介绍学科范围分析,只做了主题词和副主题词的分析举例。这种介绍主题分析方法的讲述方式看似具体,但并没有说透事物的本质。比如,为什么主题词是“骨肿瘤”,副主题词是“放射治疗”?两者能不能转换?这些不讲清楚,用户在对其他课题做主题分析时仍会不得要领。这种讲述是“一串钥匙”中的“每把钥匙”还要带“一把小钥匙”的方法。

2.3 专设章节介绍“主题分析”的方式

这种方式一般只出现在信息分类或主题标引的教材或手册中,如《中国分类主题词表教程》、《中国分类主题词表使用手册》等。检索用户培训教材中很少有专设“主题分析”章节的。但李长信老师等人主编、科学出版社出版的《医学文献检索》一书将“主题分析”作为其第三章,用整章篇幅加以介绍。其中介绍了主题类型、主题结构、主题概念提炼、主题分析方法等。这对于用户掌握如何提炼主题概念,以及将其转换成检索标识这一主题分析关键问题的方法是非常有益的。但该教材将标引工作的“主题分析”与检索时的“主题分析”混为一谈,比如其中的主题分析第一步文献审读,就忽略了用户是对课题进行分析。课题是用户自己的课题,用户只要知道怎样将反映课题内容的学科分类和主题概念语词分析出来,并转换成检索标识即可。文献审读对标引人员非常必要,因为文献作者的意图,文献内容的主题,文献的学科属性等,标引人员只能审读后才能进一步分析,并找出文献的学科分类和主题概念语词,将其转换成检索标识,正确地标引和揭示文献。尽管该教材在专

设的“主题分析”章节中存有标引培训与检索培训相混的不足,但这种方式是应该推荐的。只有专设“主题分析”章节,才能真正讲清楚“主题分析”的知识内容,让用户摸清“主题分析”的脉络,做到与标引人员建立检索系统时,对同一信息内容的主题分析结果相一致,从而提高检索效率。

3 检索用户培训中“主题分析”的内容设置

为了使检索用户真正掌握“主题分析”的知识内容,做到在检索时其主题分析结果能与标引时的主题分析结果相一致,本文提倡专设“主题分析”章节。

用户课题的主题分析是在用户的专业知识范围内,虽然也涉及检索语言,但用户接受起来效果是不同的。这种认识是基于用户对自己课题的学科隶属关系、专业术语的同义词、近义词词都比较熟悉的基础之上的。只要培训时能阐明课题专业概念语词的分解、提炼、转换以及学科分类等方法,就能使检索用户适应任何检索工具或数据库的检索要求。

在培训者设置“主题分析”的内容时,应删掉其中文献审读的内容,因为用户在检索前一般并没有相关文献,需对课题进行分析后才能查找所需的文献。基于这种认识,“主题分析”的内容设置应该包括:

(1)主题分析的基本问题。如主题分析的概念解释,什么是课题的主题,标引和检索主题分析结果趋于一致的重要性、要求、目的等。

(2)主题分析角度。包括分析角度与课题学科性质的关系;分析角度与课题需求的关系;分析角度与课题将要使用的检索标志特点的关系。

(3)课题的主题类型分析。包括主题类型的概念解释;各种主题类型(单一主题和多主题、单元主题和复合主题以及复合主题的多种类型);复合型主题转换成检索标识时应注意的问题,如“激光在医学上的应用”,其主题类型为概念联结型复合主题,“应用”为联结词。

(4)课题的主题结构分析。包括:主题结构的概念解释,即构成主题的各个要素及它们的作用和相互关系。主题结构的分析,即对复合主题类型的课题进一步分析它是由哪些主题因素构成的,各主题因素的作用以及各主题因素之间的关系,也就是对复合主题进行分解和认识的过程。复合主题的主题因素类型(主体因素、通用因素、空间因素、时间因素、文献类型因素)。主题结构分析的目的,即判断课题的中心、动态部分和限定部分,用以把握主题的主要成份和次要成份,从而对课题内容所涉及的主题概念进行提炼、精选、取舍、最终确定出检索标识。比如:“胃动脉的功能”,中心部分是“胃”,第一限定部分是“动脉”,第二限定部分是“功能”,其分类范围为内科的人体部位;主题词是“胃动脉”(如用《中国分类主题词表》有“胃动脉”这一专指词),副主题词为“功能”,因为课题的主体因素

是“胃动脉”,通用因素是“功能”,属“事物——方面”型概念限定复合主题。主题概念词的提炼、取舍与转换。其含义是将反映课题内容的、具有实质意义的词汇识别出来,目的是有助于转换成将要使用的检索工具所需要的检索标识。比如“广西近代工业发展概况”,在判断了主题类型和分析出主题结构各因素作用后,会发现其中有实质意义的词为“工业发展史”、“广西”、“近代”,而“概况”无实质意义,可以不用作检索词。若使用《中国分类主题词表》进一步规范主题词,可得“工业经济”、“经济史:近代史”两个主体因素的规范主题词和位置因素词“广西”。只要分析正确,转换成检索语词是比较容易的。能分析出课题中代表中心问题的语词和代表限定部分的语词,并从中删除无实质意义的语词,就基本上掌握了确定主、副主题词的方法。检索入口词的选择不但容易并且准确性也较高。

(5)主题分析方法。对于检索用户来讲,最合适、最容易掌握的主题分析方法是主题结构模式法。就是按照事先设计好的主题结构模式提炼相关主题要素,分析主题要素之间的关系。主题模式可表述为“主体因素——通用因素——空间因素——时间主题——文献类型因素”。课题中包含哪个因素就分析哪个因素,有的因素不只一个就要全部分析出来,以便筛选和进行匹配。比如“90年代美国政府制定毒品问题的计划”,主体因素是“毒品”、“美国政府”,通用因素是“计划”,空间因素是“美国”,时间因素是“90年代”。实际检索可以用“毒物走私”、“吸毒”、“美国”、“90年代”匹配检索文献。需要说明的是,用《中国分类主题词表》规范“毒品”一词,只能用“毒物——走私”、“吸毒”来表示。若有人查找美国政府的各种计划,再使用“美国政府”、“计划”,从隐含角度检索美国政府的计划。因标引人员从隐含角度揭示了这一隐含主题。

使用上述“主题分析”内容设置方式培训用户时,用户能否做出如下分析结果的格式即可检验培训效果。以本文曾使用过的几个课题为例:

(1)“激光在医学上的应用”

主题类型:概念联结型复合主题。

主题结构:中心问题为“激光”(主体因素)、“医学”(主

体因素);动态部分为“应用”(通用因素)。

有实质意义的语词:“激光”、“医学”、“应用”。

查表转换:主题词为“激光”、“医学”,副主题词是“治疗应用”(查表得到专指度较高的词)。

(2)“胃动脉的功能”

主题类型:概念限定型复合主题。

主题结构:中心问题是“胃”(主体因素);限定部分是“动脉”(通用因素)、“功能”(通用因素)。

有实质意义的语词:“胃”、“动脉”、“功能”。

查表转换:主题词为“胃动脉”,副主题词是“功能”。

(3)“广西近代工业发展概况”

主题类型:概念限定型主题。

主题结构:中心问题是“工业发展”(主体因素);限定部分为“广西”(位置因素)、“近代”(时间因素)。

有实质意义的语词:“工业发展史”、“广西”、“近代”。

查表转换:主题词为“工业经济”、“经济史:近代史”,副主题词为“广西”。

检索用户如果能够熟练地掌握“主题分析”的上述 5 个内容,那么检索用户与标引人员在对待同一内容的信息资料上将会思路一致。而且能了解各种检索工具或数据库编制者的初衷和编制方法,使用时会感到非常便捷。

参考文献

- 1 张琪玉. 情报语言学基础(增订第 2 版). 武汉:武汉大学出版社,1997
- 2 陈界等. 医学信息检索与利用. 北京:中国科学技术出版社,1998
- 3 李长信等. 医学文献检索. 北京:科学出版社,1995
- 4 陈树年. 中国分类主题词表标引手册. 北京:北京图书馆出版社,1998

李洪建 河北大学图书馆副研究馆员。通讯地址:保定市合作路 1 号。邮编 071002。

顾潇华 河北大学信息管理系副教授。通讯地址同上。

(来稿时间:2001-07-25)

(上接第 68 页)参考文献

- 1 周和平. 关于建设中国数字图书馆工程的问题. 中国图书馆学报,2000(5)
- 2 吴志荣. 论数字图书馆的组织模式. 图书馆杂志,1999(9)
- 3 余日泰. 数字图像图书馆建设中的主要技术. 图书情报工作,2000(8)
- 4 李玉安. 电子图书馆、数字图书馆研究与实践述评. 中国图书馆学报,1996(6)
- 5 王世伟. 论数字图书馆的特点及其对当代图书馆学教育的影响. 图书情报工作,2001(3)
- 6 赵继海. 数字图书馆发展若干领域的评析. 图书情报工作,2001(3)
- 7 赵金一. 新千年的挑战:关于中国数字图书馆建设实践

的思考. 图书情报工作,2001(3)

- 8 盛小平. 数字图书馆的知识组织. 图书情报工作,2001(3)
- 9 党跃臣,张晔,王辑华. 数字图书馆信息资源建设中的版权问题. 图书情报工作,2001(3)
- 10 范晓鹏. 关于我国数字图书馆建设的探讨. 图书情报工作,2001(3)
- 11 张菽,李素娥. 数字图书馆建设中的关键问题及其对策. 图书情报工作,2001(3)

曹学柱 秦莲霞 孟遂珍 中国气象局图书馆工作。通讯地址:北京中关村南大街 46 号。邮编 100081。

(来稿时间:2001-06-04)