

●卜书庆 陈树年

《中国分类主题词表》修订的整体构想

摘要 分析了《中国分类主题词表》修订的必要性和修订的重点,阐述了新版词表的结构系统、修订方法和程序。

关键词 情报检索语言 分类主题一体化 中国分类主题词表 词表修订

分类号 G254.242

ABSTRACT The authors analyze the necessity and the key points of the revision of *Classified Chinese Thesaurus*, and propose the structure, methodologies and procedures for the revision.

KEY WORDS Information retrieval language. Integration of classification and subject indexing.

Classified Chinese Thesaurus. Revision.

CLASS NUMBER G254.242

1 《中国分类主题词表》修订的必要性

《中国分类主题词表》自1994年出版至今已7年。它的问世,对我国图书情报机构的文献整序和编制检索工具,特别是对主题法的普及和主题检索系统的建设起到重要的推动作用。随着科学技术的发展与社会的进步,对它的修订有必要提到议事日程上来。这是因为:

(1)《中国分类主题词表》是在《中国图书馆分类法》(第3版)与《汉语主题词表》两种检索语言基础上编制的对照索引式的一体化检索语言,它应当是两种检索语言最新版本的对照索引,其中的一种检索语言如进行了修订,它也必须随之修订,否则就失去了应用的价值。《中图法》第4版已于1999年修订出版并被普遍使用,在这种情况下如不及时修订,使用《中图法》第4版的用户就无法使用《中国分类主题词表》进行一体化标引。

(2)《汉语主题词表》编于1974~1979年,此后就一直未进行完整的版本修订。20多年过去了,主题词还基本上是个老样子,这显然无法满足文献整序和检索的需求,主题法动态性高于分类法的优点也难以得到发挥。

(3)《中国分类主题词表》的编辑是一项很大的系统工程,它涉及的学科广泛,参加人员有全国40多个图书情报单位的160多位专业工作者,而整个编辑工作又基本是在手工条件下进行的,因此出现的错误也

是纷繁多样。这些错误在用户的使用过程中以及建立主题词规范数据库中陆续被发现,有必要通过版本的修订加以纠正。

(4)近几年计算机技术在图书情报机构迅速普及,传统的印刷型分类法和词表对计算机编目和计算机检索已不适应。出版机读版检索语言已是广大用户的呼声,研制出版《中国分类主题词表》机读版已十分迫切。

(5)计算机和数据库技术、网络和通信技术、数字化和多媒体技术的发展和运用,对检索系统的实用性和易用性提出了更高要求,其中之一就是自然语言检索。受控语言的自然语言化趋势已越来越明显,要想使《中国分类主题词表》适应新的技术环境、新的信息载体和新的用户需求,就必须进行全面修订。

2 《中国分类主题词表》修订工作的重点

2.1 主题词的修订

自1979年以来,《汉语主题词表》未进行重大修订,大量陈旧和无文献保证的词汇充斥其中,而大量反映新科学技术和新事物的词汇却没有收入,严重影响主题标引质量。主题词的修订应着重从以下9个方面考虑:根据文献保证的原则增补新主题词;大量增补入口词,使《中国分类主题词表》朝与自然语言结合方向迈进;修订部分原主题词相对陈旧的表述形式;把部分标引频率较低的主题词降为入口词;删除基本无文献保证的主题词,或进行归并;修订部分主

题词之间的语义关系;审查并修订词族索引;增加主题词的英译名;修订主题词字顺表的编排结构。

2.2 类目与主题词对应关系的修订

根据《中图法》第 4 版新版本,全面审查和修订主题词与类目的对应关系,同时纠正《中国分类主题词表》第 1 版的各种错误。

2.3 开发《中国分类主题词表》编辑管理系统

这是使词表的修订和维护全面实现计算机化,以及开发机读版词表的关键项目。管理系统的开发首先要解决分类法数据和主题词数据的记录格式,目前已完成《中图法》机读格式和《主题词规范格式》的设计,并以此为依据建立了相应数据库。《中国分类主题词表》可通过这两种数据格式的连接和转换,实现对分类—主题一体化数据的记录。词表管理系统应具有录入和修改、纠错、链接、自动生成各种索引、多种显示模式、编辑、排版、检索、统计等全方位的功能。

2.4 研制《中国分类主题词表》电子版

词表的电子版应满足各种条件的查询:号码的和语词的、精确的、模糊的、任意词素的。它具有满足各种需求的显示功能:不同范围的、不同详简层次的、多窗口的显示。它通过超文本技术把使用类目与交替类目、类目与参照、类号与主题词、主题词与参照项(用、代、属、分、参、族)等链接起来,实现查找过程的任意切换。它具有友好的用户界面和帮助系统;具有与标准编目系统和检索系统的接口,实现编目过程分类号和主题词的固定字段的自动录入,提高编目效率;能通过编目系统、检索系统的连接进行标引词频或检索词频的统计等。

3 新版《中国分类主题词表》的结构体系和产品形式

《中国分类主题词表》第 1 版只有印刷本,2 卷 6 册,篇幅巨大,使用不便。因此,电子版词表应是本次修订的重要内容,也是《中国分类主题词表》第 2 版的主要产品形式。考虑到部分图书情报机构在一段时期内仍编制手工检索工具,词表的印刷版形式仍应保留。

3.1 对照索引形式编制结构的不足

以我国最通用的两种检索语言为基础、以对照索引形式编制的一体化检索语言很适合我国的国情,这种基本形式应得到肯定并不断改进。但它自身的结构也存在不少问题:

(1) 过去编制《中国分类主题词表》时考虑较多的

是应用于编制手工检索工具和利用机编系统生成手工检索工具,以及为降低主题标引的难度,十分重视主题词串(即主题词组配标题形式)的编制和显示。字顺表中大量的主题词串与主题词混合排列,造成对主题词查找的噪音,而用户使用字顺表的主要目的不是查找主题词串。

(2) 在主题词对应的分类号后加标“圈码”,意在指示仿分复分的依据。但实际进行仿分或复分时,必须回到分类表中详阅类目注释,在字顺表进行标注就没多大意义。

(3) 编制之初因考虑把分类法作为主题词的等级分类索引使用,在主题词字顺表中就省略了属项和分项,只有用代项、参项和族项。分类表虽然是等级体系,但也包含着很多非等级关系,与严格属种关系的词族索引有较大差异,省略属项和分项对用户判断词义很不利。况且分类表是另一分册,来回对照翻检也相当费事。

(4) 没有英汉对照索引,对标引英文文献来说就缺少了一个辅助工具。

(5) 类目与主题词的对应关系,设计比较复杂,分为专指对应、含义对应、组配对应、单个主题词对应、主题词串对应等多种形式,增加了编制难度。有的对应方式在使用中也无太大意义。

根据这些问题和不足,分类主题词表的基本结构可根据印刷版和电子版不同的功能分别确定。

3.2 印刷版《中国分类主题词表》结构

由类目——主题词对应表(分类表)、主题词——分类号对应表(字顺表)、英汉对照索引 3 个组成部分。

(1) 分类表的款目结构:

分类号	类名	类目对应的主题词串或主题词
	注释	类目含义对应的主题词

(2) 字顺表的款目结构:

主题词	分类号
英译名	
注释项	
用代项	
属项(直接上位词)	
分项(直接下位词)	
族项(在族首词下词族全显示)	
参项	

(3) 英汉对照索引的款目结构:

英译名

主题词 分类号

上述编制结构,简化了“类目——主题词对应表”的对应形式,取消了“主题词——分类号对应表”的中组配标题款目,同时加强了对主题词语义关系的显示,使之更接近普通主题词表的模式,增加了英汉对照索引。

3.3 电子版《中国分类主题词表》结构

由分类表文档、主题词文档、词族文档、主题词组配标题文档等以及检索系统组成。类目显示和主题词显示的形式与印刷版基本相同,但可以进行属项、分项和族项的全显示。

4 词表修订的基本程序和方法(不包括系统开发)

修订工作应充分利用编辑管理系统的功能和各种已有数据库资源,避免人海战术和尽量减少手工作业的范围。人的智能劳动和手工作业应主要用在:统计数据和分析、制订选词和删词标准、预选新主题词、确定新增词语义关系和分类属性、类目与主题词的对应及组配标题编制、检查和纠正第1版中的错误等;其他作业主要由计算机和管理系统完成。

4.1 数据统计

只有通过数据统计才能准确地掌握现有主题词的标引频率,了解其文献保证率,从而判断原主题词选词的科学性和实用性,为修订提供可靠依据。也只有通过数据统计才能全面调查新学科、新事物、新主题在文献中的关键词(自然语言)表述形式,为新增主题词提供依据。数据统计可用《中国国家书目》、《中文社科报刊篇名数据库》、《中文科技期刊数据库》进行。这3个数据库基本涵盖了近10年来中文图书和期刊论文的数据,完全能调查出各学科领域标引用词的情况。数据统计须用一定的软件分别统计各数据库的标引词使用频率、标引词对应的分类号和分类号对应的标引词,然后将这3个数据库的统计数据分别加以合并,形成3种综合统计数据,作为数据分析和修订的依据。

4.2 统计分析、确定选词标准、预选新的主题词

分析标引词统计频率,初步确定标引频率(图书与期刊分别确定)与选词的关系,大概有以下几种情况:(1)标引频率极高的,一般属通用词,如设计、工艺、研究等,从中选取词表未收的通用词;(2)标引频率很高或较高的(可能是正式主题词也可能是关键词),作为主题词预选的对象;(3)有一定标引频率的,

作为入口词预选的对象;(4)标引频率很低的,作为删除、合并主题词和建立主题词用代关系的依据。在此基础上结合各种因素对选词标准进行调整,如历史文献、“文革”文献虽然某些词使用频率很低,应予保留,但统计数据中可能反映不出来。然后进行主题词的预选。

4.3 建立新的主题词规范数据库

主题词规范数据库是《中图法》编委会以“主题词规范数据格式”,在《中国分类主题词表》第1版基础上建立的。它包括主题词的全部数据,经过更新后可以作为修订词表的基础。首先,要把规范库对应的分类号改为《中图法》第4版的分类号,形成新分类号与原主题词的对应。其次,结合标引词词频统计数据和预选的主题词更新规范数据库,这一步工作量最大,基本要靠人工进行,也是词表修订关键的一步。其中包括:

(1)重新审定和调整主题词。原主题词无文献保证的(包括历史各阶段的文献),删除。原主题词标引频率相当低的,一般降为入口词。同一主题词概念,既有主题词也有关键词的,根据标引频率确定正式主题词与入口词;标引频率差不多的,以关键词为入口词。关键词标引频率较高,且无相应主题词的,入选为正式主题词(包括部分通用词)。凡是有一定标引频率的关键词,均选为入口词,同时增加其他来源的入口词。根据科学技术和社会、经济发展的趋势,有预见地增加其他来源的主题词。

(2)对新增主题词建立全面的语义关系。对正式入选的主题词、入口词建立注释、用、代、属、分、参、族语义关系,与分类号的对应以及英译名等;检查和调整原来不当的语义关系。这步工作可参考“标引词对应的分类号”统计数据进行。

(3)更新词族索引。根据对主题词修订的结果,生成一个供修订用的词族索引。主题词的增删改和新建立的语义关系,不可避免地出现许多属种关系的逻辑错误和等级链的断裂,需要经过人工检查、调整,最后完成词族索引的更新。

上述工作都需要反复调整才能完成。

4.4 建立类目——主题词对应数据库

基本步骤如下:(1)使用主题词规范数据库生成一个“分类号——主题词对照索引”。(2)使用词表管理系统把“分类号——主题词对照索引”的数据装入《中图法》第4版机读数据库,形成《类目——主题词对应表》的雏形。(3)以新的主题词规范数据为依据,为新增

●孙建军 郑建明 成 颖

面向 21 世纪的大学生信息素质教育

摘 要 开展大学生信息素质教育应从以下方面入手:落实国家的相关文件精神,把文检课作为一门正式课程;选择恰当的开课学期及教学内容;建立一支由图书馆员与专业教师相结合的信息素质课教学队伍;有充足的文献作保证,加强与系科和教师之间的联系;提高教师素质;加强现代信息技术的使用以及电子信息源的利用等方面的教学内容等。参考文献 4。

关键词 信息素质 高等教育 教育改革

分类号 G350

ABSTRACT In this paper, the author thinks that the most important factors for the information quality education of university students include: implementing related regulations of governments, establishing suitable curricula, selecting teachers with backgrounds both of librarians and teachers, collecting enough documents, etc. 4 refs.

KEY WORDS Information quality. Higher education. Educational reform.

CLASS NUMBER G350

21 世纪是社会信息化的世纪。高校大力开展信息素质教育,有其深刻的时代背景与现实意义。首先,信息素质教育的提出适应了社会发展的需要;其次,信息素质教育符合人的自身发展需要;第三,信息素质教育有助于推动全社会的信息化进程;第四,信息素质教育是信息专业教育的一大趋势^[1]。在面向 21 世纪大学生信息素质教育进行深入阐述之前,有必要对信息素质这一概念加以界定。

1 信息素质概念源流

信息素质^[2] (Information Literacy) 是一动态发展概念,在不同的发展阶段,它有着不同的内涵与外延。信息素质脱胎于图书馆素质 (Library Literacy)。图书馆素质主要包括如何利用图书馆的相关技能与知识,并以此解决问题或做出决策。随着社会信息化进程的加快,信息技术已成为现代社会生产力、竞争力和经济发展的关键因素之一,信息服务的提供已经不再局限于图书馆,各种商业运营的信息机构都在提供诸如联机数据库服务、电子邮件服务、电信服务、数据分析服务等信息服务;图书馆信息服务的手段也已

及修订的类目编制主题词组配标题;全面审查、修订原有类目对应的主题词组配标题;参照“分类号对应的标引词”统计数据,全面审查、调整类目对应的单个主题词。(4) 编制主题词对应栏的参照和说明语。(5) 根据《中图法》第 4 版新增类目与主题词对应的需要,提出须再增补的新主题词。(6) 根据类目—主题词对应的结果,纠正《中图法》第 4 版明显的错误。(7) 将上述 3~6 的修订数据返回主题词规范数据库,对该数据库进行第 2 次修订;然后再重复本项的修订过程,直至从语义上、逻辑上没有错误为止。

最后是使用词表管理系统完成《中国分类主题词表》各版本编排和编写《中国分类主题词表》标引手

册。

《中图法》编委会准备用一年半到两年的时间完成修订任务。我们提出上述初步构想旨在引起图书情报界同仁对此项工程的关注,对修订的各种问题进行深入研究,为《中国分类主题词表》的修订奠定良好的基础。

卜书庆 副研究馆员。通讯地址:北京白石桥路 39 号,国家图书馆。邮编 100081。

陈树年 华东理工大学石油化工学院研究馆员。通讯地址:上海市金山卫。邮编 201512。

(来稿时间:2000-03-28)