

索引数据库质量评价标准的探讨

王梅 舒明全

(中国科学评价研究中心,武汉 430072)

【摘要】 在简要说明建立索引数据库质量评价标准的必要性和紧迫性的基础上,结合国内外现有的索引质量标准,提出了我国索引数据库的质量评价基准,并以全国高校专题特色数据库的质量评价体系为参考构建了我国索引数据库的质量评价标准的内容(包括通用标准和专用标准),最后就索引数据库质量评价问题提出了自己的续想。

【关键词】 索引数据库; 质量评价; 标准

【中图分类号】 Z89; G644

Study on the Quality Evaluation Standard for Chinese Index Data

WANG Mei SHU Mingquan

【Abstract】 This paper briefly states the importance of establishing evaluation standard for index data quality, bring forward evaluation norm for Chinese index data based on the criteria for index quality at home and abroad and set up evaluation standard for Chinese index data quality referencing the quality evaluation criteria for CALIS subject data.

【Key words】 index data; quality evaluation; standard

目前我国索引数据库已有千余种,但实际上能公开利用的并不多,大部分是自用索引数据库,这就造成成本与利益的不均衡,使索引数据库出现“断粮”现象,难以持续发展。由于索引数据库建设的投入成本高,特别需要大量专业人才坚持不懈的长期努力,才能寻求数据库的发展机会。虽然国家进行信息化建设,经常投资建立各种索引数据库,但苦于没有规范的验收标准,无法衡量索引数据库评价的合理、公正,使数据库项目鉴定夭折。索引最常见的形式是书后索引、图书索引和论文索引,从实现技术平台来说,三类索引都可以利用数据库管理,但本文主要是针对论文索引数据库来讨论数据库的质量评价问题。

1 索引数据库的质量评价基准

索引数据库质量评价目前还没有专门的标准或评价指标,但在国内外的索引标准中已有对索引质量的要求,索引质量的要求应该是索引数据库质量评价标准。以下分析了索引的国际标准、美国标准、英国标准和台湾标准中的索引质量控制部分,并提出建设中国索引质量标准。

1.1 现有索引质量标准的浏览

综观现有的国内外索引标准,索引质量是索引的基本保障。在国际标准《情报工作和文献工作——索引的内容、组织和表示准则》第六部分明确规定索引的质量控制,提出了索引质量的要求,索引篇幅的规定,强调标引的一致性,处理索引员与作者、出版商的关系^[1]。美国标准《图书馆学、情报学及出版工作》在第三部分索引的作用和一般性质内,规定索引质量的要求,提出索引的准确性要求,用户语言的适用性^[2]。英国标准《图书、期刊及其它索引的编制》在第四部分索引的类型与特征内,规定索引质量首先要适合用户需求,其次要求公正、客观地确认索引词,实现标引的深度,同时恰当处理索引员与作者、出版商的关系^[3]。台湾《索引编制标准》在第六部分的索引政策中,提出了索引质量标准的完整性、精确性与一致性^[4]。

确定索引质量控制规范,才能保证索引数据库质量。由于我国目前还没有索引标准,为了促进索引数据库建设走进规范、互用和共享发展,需要建立中国索引质量标准。

1.2 中国索引质量标准的建立

根据现有的索引标准,为了确保索引数据库质

量,我国迫切需要索引质量的基准。中国索引学会申报了《中国索引标准》编制项目,我们学校非常幸运,有机会参加《中国索引标准》起草工作,负责索引质量部分。起草内容包括索引的质量要求、标引的一致性、索引员与著者、出版商和用户问题。

1.3 索引数据库质量标准的建立

目前我们了解到的是各种索引质量标准的总则,作为索引数据库的质量标准是有宏观指导意义的,但作为具体评价索引数据库的质量可能还不够充分。我们希望中国索引标准的总则编辑后,能再编制不同索引类型的具体标准,为索引数据库建立、验收和推广提供可靠的尺度。

2 索引数据库质量评价标准的内容

索引数据库质量评价标准涉及的内容很多,我们分为通用标准和专用标准两部分。专用标准有分为印刷性索引和数据库索引的评价。

2.1 索引数据库质量评价的通用标准

借鉴现有的各种索引标准,质量评价通用标准考虑的基点是与国际标准接轨,需要满足中国用户使用文献的习惯,同时需要适应中国文献和语言的特性,比较容易理解和推广应用。经过对已有的索引标准质量控制内容分析,初步拟订中国索引标准的质量控制部分内容如下。

2.1.1 索引的质量要求

(1)一部有效的索引必须能满足用户检索文献中信息的需求。

(2)索引款目应当考虑到用户可能采用的各种不同检索途径。

(3)索引员应当客观、公正地选择主题素材和选取主题词。

(4)提高索引员的素质是保证索引质量的前提。

2.1.2 标引的一致性

(1)一部索引的标引深度、风格和版面形式应当前后一致。

(2)索引应当按照一种合乎逻辑的、均衡的、一致的和易于识别的模式进行编制。其一致性适用于主标目、副标目、倒置标目的采用,适用于既定的标引深度、术语、细分、倒置、交互参照、出处项的使用以及总体风格和版面形式。

(3)采取以下方法,提高索引的一致性:

确立并遵循科学、实用的基本标引规则和标引工作程序;

采用权威的标引工具,如字典、主题词表等,

或咨询有关专家;

制定索引标引工作细则和记录标引的异动记录;

同一部索引的编制工作由两个或多个索引员完成时,应认真协调。

编制索引时仔细检查、修改可以确保索引的标目和副标目选用一致,准确设置交互参照、出处项,正确使用标点符号和空格,必要时用附加款目取代交互参照。

2.1.3 索引员与著者、出版商和用户

如果索引员能够与著者一起就被标引文献进行讨论,了解出版商的要求,并充分考虑索引潜在用户的要求,将有利于提高索引的质量和一致性。

出版商如果有任何特殊的规定或要求,必须提前通知索引员。例如,在标目和出处项之间是否必须加逗号,是否对索引的页数有要求,是否对所用的排版系统有特别的限制。不应当向索引员编制索引工作施加不必要的约束,而应当给予索引员圆满完成索引编制工作所需的足够的信息和指导。

如果在著者、出版商、数据库管理员和索引员之间因为诸如编制成本、时间要求、版面设计、索引篇幅等问题发生利益冲突,那么,用户利用索引有效查找信息的期望应该是要考虑的首要前提。

中国索引标准(总则)的质量控制内容可以作为索引数据库的质量评价通用标准。因为索引数据库其本质是与印刷型索引是一致的,只是在处理手段和利用方式上优于传统索引,笔者认为中国索引标准(总则)的质量控制是进行索引数据库的质量评价的基本理念。

2.2 索引数据库质量评价的专用指标

虽然确定了索引数据库质量评价通用标准,但由于质量评价通用标准是多种类型共用的准则,无法判断各类索引的特殊属性,可操作性较差。因此,需要制定索引数据库质量评价的专用指标。

(1)现有索引质量评价的专用指标

1994年《论索引质量评价的指标体系》一文针对图书索引和连续出版物索引,提出了索引质量评价的指标体系,采用二级指标结构,列举了10类评价指标。评价内容包括索引编制原则、收录范围、编制方法、款目结构和排印,评价指标内容包括索引类型、使用说明、密度、范围、规范性、款目结构、注释与倒置、刊龄、频率与及时性、检索、排印^[5]。这些专用指标比较适合印刷型的论文索引,同时也为索引数据库的质量评价指标提供了借鉴。

(2)现有文献数据库质量评价的专用指标

上文同时提出了文献数据库质量评价指标体系。评价指标体系内容包括类型、词表、排列、检索、准确性五类。二级指标:索引类型包括作者索引、主题索引、篇名索引;标引词类型包括叙词、主题词、关键词;词表包括新词扩充、词表更新频率、自动分词;排列包括顺排和倒排,单处理和批处理;检索包括单元词与多元词,截词与替代词,主题词与叙词;准确性包括主题词标引、文献出处、款目指向数^[5]这种文献数据库质量评价指标体系可以认为是索引质量评价的专用指标雏形。

(3) 索引数据库质量评价的专用指标

质量评价体系

索引数据库质量评价的专用指标应涉及组织管理、数据库建设、数据库服务和技术平台。详细内容见图1。索引数据库质量评价的专用指标应在质量评价体系的基础上进行细化,分别列出二级指标内容。同时对二级指标内容确定测评的度量级别,可以采取分值计算方法。详细内容见表1。

质量度量的过程

索引数据库质量评价是按一个完整的评价行为过程或判断问题的解决过程来展开的。这个过程由多个紧密相扣的环节组成,各个环节对评价主体在认识、能力、知识和技术等方面的要求,就构成了评价的内涵,而这个内涵的灵活缩放,即可验证质量评价的



图1 索引数据库质量评价体系

目标和内容。这个评价行为过程由5个阶段组成:明确评价需求;收集评价数据;判断评价数据;总结评价结果;评价过程和结果。

4 个例分析

全国高校专题特色数据库是CALIS“十五”建设的子项目之一,目标是建成一批具有中国特色、地方特色、高等教育特色和资源特色、服务于高校教学科研和国民经济建设、方便实用、技术先进的专题文献数据库。已有61所高校的75个项目获立项。由于特色数据库子项目是分散建设,质量评价显得格外重要,在项目申报、评审立项、项目实施、中期检查、试运行检查、成果验收六大环节需要过程质量评价标准,特别是项目的中期检查和终期验收都需要更详细的评价指标。专题特色库质量评价指标如表1所示。

表1 专题特色库质量评价指标

评价项目	评价内容		分值
项目管理 (15分)	自检报告	提交项目自检报告	10分
	人员配备	1、有固定的建库人员	1分
		2、有学科专业人员参与建库	1分
	工作文件	1、项目管理文件	1.5分
		2、业务文件	1.5分
数据加工 (30分)	数据格式	数据加工格式符合规范要求	6分
	数据规模	数据库应有一定规模的数据量,新增数据应为申报立项时数据量的50%	9分
	文献类型	1、数据库建库内容应不少于三种文献类型	6分
		2、全文数据与二次文献比例不少于20%	3分
	质量控制	1、对象数据误链率	2分
		2、文本错别字和断句差错率	2分
		3、图象倾斜、不清晰率	2分
元数据标引 (20分)	规范应用	1、是否采用了项目组规定的元数据规范	3分
		2、项目组规定的编码体系是否采用	3分
		3、元数据扩展是否遵守扩展规则	2分
	标引质量	1、项目组要求的必备字段是否具备	3分
		2、著录是否严格遵守了著录规则	3分
		3、编码体系的著录是否符合要求	3分
		4、著录详细程度	3分

续表1

评价项目	评价内容		分值
服务功能 (20分)	检索方式	数据库应能提供一定的检索功能	5分
	检索途径	数据库应能提供基于题名、作者、关键词、日期等检索途径	5分
	服务方式	1、提供数据库信息发布	3分
		2、其他服务,如提供交互功能、文献传递	3分
	统计功能	1、提供访问统计或点击数统计	2分
		2、提供用户分析报告	2分
系统平台 (15分)	数据安全与备份	是否具有数据备份与恢复功能及工作机制	4分
	软、硬件设施	1、硬件设施:是否有独立的服务器	4分
		2、系统软件平台:是否采用通过Calis认证的本地系统	4分
		3、网络通讯条件:与教育网的网络带宽(10M/100M/其它)	3分
	100分		100分

3 索引数据库的质量评价续想

索引数据库是集用户需求、文献资源、数据处理和技术平台为一体的系统,质量会出现在不同的方面;索引数据库是集多专家的集约式作业方式的智力活动,众多专家的思维更难保持协调一致;建立索引数据库的质量评价体系越来越必要。虽然本文尝试探讨索引数据库的质量评价问题,提出我们的思路,但建立索引数据库的质量评价指标体系、控制政策、度量方式和质量保障还需要更多专家的研究与实践。

参考文献

1 情报工作和文献工作——索引的内容、组织和表示准则

(国际标准).索引技术和索引标准.北京:北京图书馆出版社,1997

- 2 美国国家标准委员会.图书馆学、情报学及出版工作——索引的基本标准. ANSI Z39.4-1984.索引技术和索引标准.北京:北京图书馆出版社,1997
- 3 英国标准.图书、期刊及其其它索引的编制.索引技术和索引标准.北京:北京图书馆出版社,1997
- 4 索引编制标准(台湾).索引技术和索引标准.北京:北京图书馆出版社,1997
- 5 王辉.论索引质量评价的指标体系.图书与情报.1994(3):53-56

作者简介

王梅,女,武汉大学图书馆研究馆员。

舒明全,男,硕士,武汉大学图书馆馆员。

全球信息技术竞争力排行榜

总部设在瑞士日内瓦的世界经济论坛发布的最新全球信息技术竞争力排行榜显示,美国在2004年的排名跌至第五位后,2005年从新加坡手中重新夺回榜首位置。

据《金融时报》28日报道,世界经济论坛在对全球115个经济体进行调查后公布了这份《2005年度全球信息技术报告》。新加坡、丹麦、冰岛和芬兰分列第二至第五位。

报告说,美国荣登榜首是因为拥有完善的硬件设施和积极的市场环境,同时企业和政府在新技术使用方面也达到较高水平。报告还强调,信息通信技术已经成为经济增长和竞争力增强的一个重要推动力。在欧盟成员国中,英国、德国和法国在信息技术领域最具竞争力,在排行中分列第10、第17和第22位。

世界经济论坛对全球信息技术竞争力的排名依据的是各经济体的法律规范、基础设施等宏观信息技术环境、人民的受教育程度以及企业和政府在信息技术研发方面的投入等因素。

摘自《世界数据报告》<http://data.icxo.com/htmlnews/2006/03/30/829571.htm>[2006-07-18]