

常用中医古籍数据库评价与分析

★ 李兵 (中国中医科学院2006级硕士研究生 北京100700)

摘要:在对中医古籍数据库的特点和用户对象进行调查的基础上,提出了中医古籍数据库的评价要素和指标,进而对中医古籍数据库评价体系的构建进行探讨,并就此对常用的中医古籍数据库进行了对比分析。

关键词:中医古籍;评价指标;数据库

中图分类号:R 22 **文献标识码:**A

经过20年的发展,古籍数据库的建设取得很大进展。据统计,目前已有的中文古籍数据库近百个,大型的网络古籍数据库也是可圈可点。中医古籍数据库和大型古籍数据库中的中医药部分是古籍数据库的重要方面。对中医古籍数据库进行评价可以从整体的角度指导数据库的建设和改进。在实践层面上,可以为中医古籍数据库的质量控制提供帮助,并为用户在利用方面提供指导。建立一套合理、可行的古籍数据库评价指标体系是必要的。

1 中医古籍数据库建设概述

中医古籍数据库的建设方兴未艾,大学图书馆、科研院所及一些商业单位都参与其中,也取得了一定的成果,如《中华医典》目前已收录中医古籍近千种;龙语瀚堂典籍数据库中有专门的中医药文献库;《文渊阁四库全书电子版》中设立了中医药专版;爱如生系列数据库也推出了医书集成产品;金图国际中医药古籍资料库、国学宝典等数据库中都含有大量的中医药书籍,此外,即将投入使用的中医古籍资源数据库已收录中医古籍850余种,在建的中医药古代文献知识库也已收录了大量中医古籍。可以说,通过对中医古籍的数字化建立中医古籍数据库已经是中医古籍整理和利用的新的方向。

2 中医古籍数据库的评价指标

要对众多的中医古籍数据库进行评价就要首先确立评价指标体系。由于古籍数据库的评价工作没有前人的经验可以借鉴,我们在中医古籍数据库建设实践和用户调查分析的基础上,参考网络数据库评价指标提出针对中医古籍数据库的评价指标体系。结合用户调查并综合分析诸多网络文献数据

库,确立了内容、检索功能、辅助功能和服务功能四个要素作为评价标准的一级指标。

2.1 内容评价 中医古籍数据库是实用性的文献数据库,其内容可以说是数据库的生命,同时,数据的质量和易用性也是影响用户对数据库评价的重要因素。内容评价应包含收录中医古籍的种类和数量、珍善本所占比例、著录标准和标引方式、全文数据的质量、图像文本格式等二级指标。

2.2 检索功能 快捷准确的检索功能是古籍数据库应用的一大优势,对检索功能的评价主要考察其是否为用户方便、准确地查找信息提供多途径支持。其二级指标应包括检索入口方式;检索方式是否具有多途径检索;检索技术是否支持全文检索、算符检索、关联检索、模糊检索等;检索结果的精准度和显示方式如何;是否具有检索工具等。由于检索功能涉及的要素较多,亦可在二级指标下设立三级指标。^[1]

2.3 辅助功能 古籍文献的阅读、研究往往需要一些参考书和研究工具,古籍数据库能否为用户提供相应的辅助支持功能也是衡量一个古籍数据库质量的一方面。辅助功能评价的二级指标包括是否提供阅读帮助工具,如放大镜、书签、注释工具等;是否有相应内容的参考工具,如历史年表、纪年转换、参考文献等;是否具有系统统计功能,如词频统计、系统统计等。

2.4 服务功能 古籍数据库建设的目的是为用户服务、提高古籍的利用度。服务功能的评价实际上是反映古籍数据库能够为用户提供方便的水平,完善的服务功能是古籍数据库建设过程中要考虑的因

素。关系到服务水平的指标包括如下方面:一、界面设计和个性化服务的提供;二、数据的获取方式,是以网络形式还是客户端形式,共享方式和收费情况如何,是否允许数据的复制和打印;三、用户管理功能,系统维护和更新情况;四、互动交流平台。

3 中医古籍数据库评价体系的构建

由于建立中医古籍数据库的评价体系没有相关工作基础,参考现代信息评价方法,可以采用文献调研法、网上特菲尔专家调查法和基于指数标度的层次分析法。^[2]具体过程如下:

3.1 减轮特菲尔专家调查

3.1.1 专家选择 按照特菲尔法的要求,选择本领域相关的专家 10~20 人,在选择过程中应兼顾各专业专家的分配。所选专家应在各类型机构中占一定比例,还应涉及中医学、图书馆学和计算机各专业,以保证指标确定的合理性。

3.1.2 调查表设计 按照上述评价指标设计专业的调查表,交由专家进行评审。评审应分次进行,先对一级指标进行调查,然后是二级指标。各指标按“重要”、“较重要”、“一般”、“不重要”给出四个评

分等级,分别赋以相应分数。

3.1.3 专家评审 每位专家对调查表各项指标打分,分别计算一级指标和二级指标的得分情况,统计各项指标在总分中所占比例,最终确定指标组成和权重。

3.2 层次分析法

3.2.1 确定评价指标的权重 层次分析法(AHP)在解决多目标决策问题方面具有比其他方法简便实用的特点,可用于计算获得各项指标的权重。^[3]在计算各指标满分频度之后,通过数据处理计算标度,用指数标度构造判断矩阵,计算过程可由 java 程序实现。

3.2.2 一致性检验 由于事物的复杂性或认识的片面性,对判断矩阵求出的值可能出现不合理现象,因此要对所得权重进行一致性检验,若不满意,则需重新调整矩阵。

4 常用中医古籍数据库的对比分析

基于上述评价指标体系,结合目前中医古籍数据库的使用情况,本文对常用的古籍数据库进行对比。见表 1。

表 1 常用中医古籍数据库的对比分析

指标	数据库				
	中华医典 ^[4]	龙语瀚堂典籍数据库 ^[5]	书同文系列之中医类	爱如生医书集成 ^[6]	
内容要素	收录内容	民国以前 历代医学古籍 1000 部	历代中医古籍 258 种	中医学经典著作 104 种	历代中医名著 500 种, 多为善本
	著录标引	列有目录索引	目录、标题标引	目录索引	ASM 分类法著录
	数据质量	有一定的差错率	采用四字节 编码技术, 差错率在 1/10000 以内	OCR 识别辅以人工校对, 差错率在 1/1000 以内	有一定的差错率
检索功能	数据格式	文本数据、无原文图像	图文对照、图像为 IMG 格式	图文对照黑白图像	文本 ABT 格式、黑白影像
	检索入口	单一	多入口	多入口	多入口
	检索方式	全文检索	多途径检索	全文检索	多途径检索
	检索技术	全文检索	生僻字检索、精确检索、智能检索	组合检索、限定检索、关联检索	组合检索、逻辑检索、关联检索
	检索结果	结果全, 噪音大	结果全, 噪音大	有一定的漏检和噪音	ASE 古籍专用检索系统, 结果较为精确
辅助功能	阅读工具	翻页	翻页、缩放	翻页、放大、笔记、书签、标点、标记、手写输入等	翻页、缩放、标点批注、版式设定
	参考工具	辞典、计量换算	字典、文字、音韵、训诂等小学工具	有联机字典、纪年换算、参考文献	常用字典、版本对照、繁简转换等
	系统统计	无	无	词频统计	无
服务功能	界面设计	友好、形式单一	一般、形式单一	友好、可进行个性化设置	友好、可进行个性化设置
	获取方式	光盘版、限字复制	网络版、可复制、打印	光盘版、可复制	单机版、可编辑下载、打印
	互动交流	在线帮助	使用帮助、意见反馈、在线交流	使用帮助	纠错、扩充机制

证型标准化研究的文献统计分析方法学探讨

★ 刘艳 陈丽云 卫洪昌 (上海中医药大学基础医学院 上海 201203)

摘要:初步探讨了基于现代文献统计分析方法上的证型研究现状,基于复合证、单证、症候要素、证—症对应关系等四种研究角度,从频次分析、病例数构成比分析、证—症对应关系分析等三种研究方法对证型标准化研究的文献统计分析方法学进行了探讨。认为文献分析方法可作为中医证型标准化研究的一种有效的辅助手段,但其结论的准确性依赖于文献资料的准确性和研究者对数据处理的准确性。

关键词:文献统计分析方法;证型标准化研究

中图分类号:R 241 **文献标识码:**A

Discussion on Methodology of Literature Analysis on Research of Pattern Standardization

LIU Yan, CHEN Li-yun, WEI Hong-chang

Basic Medical College, Shanghai University of Traditional Chinese Medicine Shanghai 201203

Abstract: This paper approached the methodology of literature analysis which is widely used recently in the study of standardization of syndrome pattern of TCM. Three research means including frequency analysis, analysis of constituent ratio of case number and analysis of corresponding relationship between syndrome and symptom are approached in this paper. The author argued that the method of literature analysis could be used as an effective assistant device in the study of standardization of syndrome pattern of TCM. However, the accuracy of the conclusions by this method depends on whether the included information and data processing are accurate.

Key Words: Method of Literature Analysis; Study of Standardization of Syndrome Pattern of TCM

中医标准化是中医走向世界,被更多的人接受和应用的重要一步,近十多年来,有很多学者做了这方面的研究。中医证的规范化、标准化研究是中医

标准化研究中一个十分重要的课题。要了解某种病的中医证型或征候分布规律,进而对其进行规范化,对该病进行临床流行病学调查是现代研究中公认的

5 结语

上述对中医古籍数据库的探讨和对比可能存在有待商榷的地方,但我们还是可以看出当前的中医古籍数据库还需要更加完善和发展,需要更进一步的建设和完善。良好的评价指标体系有利于规范和指导中医古籍数据库的建设,保障数据库建设的质量,规划合理的中医古籍数字化方案。在信息社会的今天,中医古籍数据库的建设应该有更高的要求,这样才能更好地为中医药事业服务。

参考文献

[1] 刘霞,刘颖.数据库检索系统的评价与选择[J].现代图书情报技

术 2004,(6):49-52.

[2] 朱庆华.信息分析:基础、方法及应用[M].北京:科学出版社,2004:74-88.

[3] 吕跃进.决策科学理论与方法[M].北京:海洋出版社,2001:50-58.

[4] 中华医典[EB/OL].[2008-12-15].<http://www.tcmbook.cn/>.

[5] 龙语瀚堂典籍数据库简介[EB/OL].[2008-12-16].<http://public.dragoninfo.cn/Introduction.aspx>.

[6] 医书集成产品介绍[EB/OL].[2008-12-19].<http://www.er07.com/index.jsp>.

(收稿日期:2009-02-09)