

中文循证医学数据库CCED的研发

张晓梅 刘传和

[摘要] 本文介绍了中文循证医学数据库研发的全过程,并对研发过程中必须注意的重要问题和系统信息服务功能的进一步完善进行了讨论。

[关键词] 循证医学;临床证据;数据库;系统研发;文献检索

[Abstract] This paper introduces the whole procedure of the development of the Chinese Clinical Evidence Database, as well as some important matter during the development and the further improvement of the information services of the system.

[Keywords] Evidence-based medicine; Clinical evidence; Database; System development; Literature retrieval

1 循证医学和循证医学数据库

1.1 循证医学的概念和意义

循证医学(Evidence-based medicine, EBM)是一种以临床证据(Clinical evidence)为基础的医学模式。它运用临床流行病学中的科学原则和评价方法,对来自临床实践中的各种证据进行系统科学的评价,去粗取精、去伪存真,获取最佳证据,并将其运用于临床决策和临床研究工作中,指导临床实践活动。

在循证医学模式下,要求任何临床实践活动,包括任何的临床研究和临床预测决策等,都必须以科学证据为主要依据,而不是依靠医务工作者的个体经验或个案先例。由此可见,循证医学模式是对以传统经验为基础的传统医学模式的变革。这种模式不仅对临床医学实践活动提出了新的要求;同时也要求要有一支专业队伍专门致力于为临床工作提供科学证据。因此,科学证据的获取和提供,就成为循证医学模式下医学信息工作者义不容辞的重要工作内容之一,循证医学中心、各种专业机构和循证医学数据库也应运而生。1997年华西医科大学成立了我国首个循证医学中心,1999年被国际Cochrane协作网正式批准注册为中国循证医学中心。目前,临床医学正处在从经验医学到循证医学模式的转变过渡期^[1]。

1.2 循证医学数据库

简单地讲,循证医学数据库是以临床证据数据为核心的数据库检索系统。循证医学数据库的研发,旨在建立一个完整的临床证据数据获取和利用体系,最终目的是要实现以科学的方法收集和整理临床医学证据,通过专业人员运用专业技术手段对证据进行加工、整理、分析、评价,并将其有机地组织在一个数据库体系中,并为临床工作者和研究人员提供获取和利用最佳临床证据的平台,为临床医学实践活动的开展提供最可靠、最方便的信息,为医疗卫生管理者和决策机构提供卫生决策依据,也为病患人员提供他们最为关心的医疗信息。由此可见,循证医学数据库的研发具有重要的现实意义。

2 中文循证医学数据库(CCED)的研发

中文循证医学数据库(Chinese clinical evidence database, CCED)以临床证据为核心,整个系统具有极强的专业性,因而给开发工作带来诸多困难。相对而言,整个开发过程中的难点不在于检索系统的设计与开发,而在于原始数据的获取和对证据数据的深加工。因此,研发工作自始至终要紧密围绕循证医学的专业需求这一核心:一方面,必须严格按照流行病学的科

学方法采集、加工和组织系统中的各类数据,这样才能保证系统数据的科学性和可靠性;另一方面,系统研发各个环节所用到的软件工具以及最终的用户系统的设计,都必须符合临床证据数据加工和检索的专业需求,突出系统数据的临床特征。

2.1 中文循证医学数据库的功能和系统结构

中文循证医学数据库系统大致包括临床证据库、加工工作库、系统管理库、检索系统和辅助信息服务系统五大组成部分(图1)。临床证据库的主要内容是原始证据和二次证据数据,加工工作库是用于证据采集和标引加工的临时数据库,系统管理库是为实现系统的管理和其它功能而建立的系统数据库,检索系统和辅助信息服务系统是面向用户开发的应用系统,内容包括系统提供的多层次多模式的检索功能和各种信息服务功能。

2.2 研发工作的步骤和主要内容

中文循证医学数据库CCED的研发工作大致包括几个方面的内容:

2.2.1 证据的收集和整理

首先,要确定证据数据的收集范围、类型和数据源。通过对馆藏的各类可利用资源(包括纸本书、刊以及电子版的各种数据库资源等)的广泛调研分析,确定中文循证医学数据库的数据采集范围和所要采集的证据类型(包括原始证据和二次证据),其中二次证据是证据库的重点内容^[2]。证据库的收录范围和结构如图2所示(图中虚框部分本次研发暂时不做)。

其次,针对选定的可用资源和范围,确定数据采集的原则和各类证据数据的采集策略(包括基本采集策略和扩展采集策略),明确具体的采集方法以及数据采集需要使用的工具^{[2][3]}。一方面,为保证证据采集的查全率和查准率,应分别提供不同的检索策略;另一方面,由于证据数据的采集检索策略非常复杂,普通的检索系统往往不能支撑复杂的检索表达式,难以检出满足条件的证据。因此,在没有现成可用的检索工具时,还需要专门为此开发专用的采集和标引加工工具。

第三,是要明确数据质量监控措施等。对于采集到的原始数据,往往与系统目标数据存在一定的偏差,必须由专业人员进行对采集到的数据质量进行监控,一旦数据质量不符合要求,必须对检索策略进行相应的调整。

2.2.2 证据原始数据的预处理和证据数据的标引

循证医学的证据数据分为不同的类型和不同的研究方面和研究内容。由于不同类型和不同研究方面的证据分别由不同的专业人员、不同的检索策略进行采集,因而采集到的数据必

作者单位:100039 解放军医学图书馆(张晓梅 刘传和)

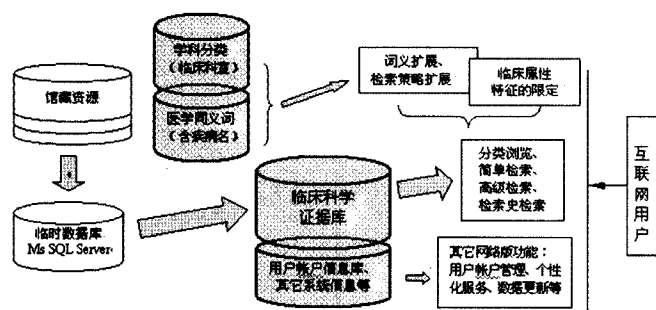


图1 系统功能和结构示意图

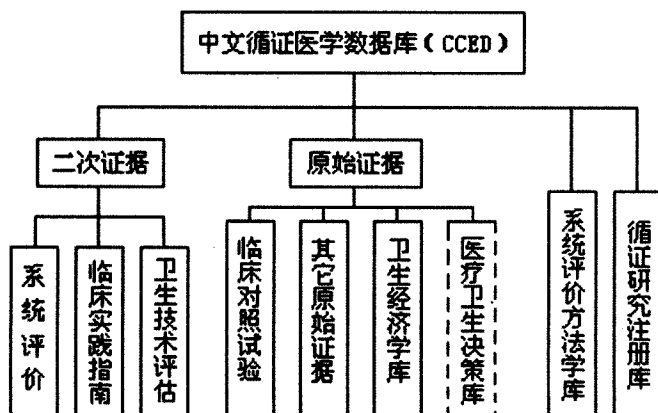


图2 循证医学证据库的收录范围和结构

然会有重复或交叉。因此，必须对收集到的原始数据先建立一个临时数据库(中间库)，对各类型的原始数据进行去重、分类标引，并对标引信息进行合并。除证据数据的类型标引外，系统证据数据的其他临床专业属性也需要进行人工或机器标引。因此，可将原始证据数据装入临时数据库，开发专门的软件工具，在完成上述各项标引工作的同时，提供人工过滤、标引和质量监控的窗口，并为最终的循证医学数据库系统输出正式的证据数据。

在此步骤中，临时数据库的结构和软件功能的设计，不仅要考虑到标引加工的工作流程和内容的需求，还需要考虑到最终循证医学数据库系统功能的需求和最终系统数据库结构，以确保加工出的数据的可靠性和可用性。

2.2.3 多模式多层次的证据检索系统平台的功能设计

首先要根据系统的定位，确定系统的主要用户群体的需求。在此基础上，突出主要用户群并兼顾其它不同类型用户的需求的需求，有针对性地确定系统的各项相应功能^[4]。系统的主要功能包括两大部分：

一是证据分类浏览：根据不同的用户，提供适应多种用户的多种分类方式，包括按证据类型分类、按临床研究内容分类、按学科分类、按临床科室设置和按疾病分类等多种模式。

二是证据检索功能：包括一些检索系统通用的检索功能，如证据的简单检索、高级检索(生成复杂的检索表达式或检索条件的逻辑运算组合等)、同义词词义扩展检索、附加临床特征限定检索、检索史记忆及逻辑运算检索等多种辅助检索手段，适应多种类型用户的多种不同检索需求。

此外，系统还提供多种格式的输出模式和简单的统计功

能。对于不同版本的系统还将根据可行性提供一些辅助的系统增值信息服务功能。

2.2.4 系统的其他功能的设计

主要包括系统其他信息服务功能和辅助系统管理功能的设计。内容主要有系统的版本设计(如光盘版或是互联网版、单机版或是网络版等)；安装、注册控制；用户合法性验证(如IP控制或其他方式的使用范围控制模式)；系统和数据的加密方式；出于系统数据安全的考虑对用户使用作出的限定等(如下载的最大记录数、获取证据文献全文的权限、下载全部或部分字段信息以及打印等的权限等)；系统数据的发布及数据更新安装相关设计；系统的增值信息服务功能设计及其他辅助功能设计(如全文服务、提醒服务、循证医学资源导航、证据数据统计分析、业界新闻报道、循证医学学术论坛等)。

2.2.5 系统的结构设计

是指对循证医学数据库系统的库结构、数据表结构及数据表间的关系、数据流向、系统功能及工作流程、程序模块结构和接口关系和接口定义等内容进行设计和定义。如确定各类证据数据的库结构；不同类型的证据是单独建立数据表还是集中在一张表中；各个数据表的结构；确定哪些功能是面向用户的，哪些功能是对用户透明的等。

2.2.6 系统的实现

系统实现主要包括系统开发环境的搭建、系统的详细设计、系统程序模块的编码和测试。在系统研发过程中，对于数据加工的临时数据库和最终的证据数据库，可以根据开发的环境条件，选用相同或分别选择不同的数据库平台和开发平台。

根据我馆数据库生产环境的实际情况，数据加工使用的临时数据库我们选用了Microsoft SQL Server 2005作为后台数据库，并利用Visual Studio 2005开发平台来开发加工工具。对于最终的中文循证医学数据库系统，后台数据库可采用SQL Server，也可以利用其他的专用数据库平台，具体要视最终产品的版本情况而定。例如，对于光盘版的中文循证医学数据库，可采用第三方专用数据库平台进行二次开发，对于互联网版的产品，则可选择通用平台，以便更好实现一些增值的信息服务功能，使用户的使用更为方便。

3 系统设计中需要注意的若干问题

循证医学数据库是以临床医学证据数据和证据的利用为核心的。无论是从系统的数据库结构和系统功能设计上看，都有着区别于普通文献数据库的显著特征。因此，在研发过程中尤其要注意以下各方面的问题。

3.1 证据库结构的设计要充分考虑到循证医学数据自身的特点

从循证医学对证据类型的定义来讲，证据数据既包括原始证据，也包括对原始证据进行科学分析评价和总结得出的二次证据。从其来源讲，可能来自专业期刊发表的文献，也可能来自权威机构或专业机构发布的文件资料、出版的书籍等。与普通的期刊文献数据相比较，循证医学库中收录的证据数据，不仅可能具有普通文献数据的题名、作者、出处(刊名/年/卷/期/页)等一系列文献数据的基本属性，更重要的是还具有作为证据的特有属性。例如，对于系统评价、临床实践指南、卫生技术评估等临床实践者最为需要的二次证据文献，用户最为关心的可能是其研究机

构、发布时间、适用范围等的一些特征,而不是文献数据的属性特征。又如,对于临床试验和随机对照试验,用户最为关心的可能是一组对于试验的特征进行描述的特有属性,如样本数量、试验原理和方法、统计分析方法等。再比如,对于患者而言,他的需求可能五花八门。检索到一条证据文献,他所关心的可能是本条证据涉及的患者的属性特征(如适用年龄段、性别等),也可能是治疗方法的副作用、或某种方法的治愈率、治疗费用等。需求决定系统的结构和设计。以上特征应在数据的采集加工和系统设计等各个环节上有所体现^[5]。

因此,如果所采集的数据来自期刊文献,则要增加新属性的标引。在可能的条件下,除了进行循证医学分类信息的标引外,还应将与临床实践相关的用户最为关心的特征属性进行标引,以突出其临床特征,这样才能从功能设计上突出临床特征,吸引不同用户,为用户提供最大方便。这也应该是循证医学数据库系统与普通文献检索系统的重要区别。

3.2 数据库的设计要充分考虑到数据采集和加工的可行性

由于循证医学数据库系统最终目的是用于临床工作的指导或参考,因此必须确保其数据是有参考价值的、真实的;结论是具有可靠性、权威性和指导意义的。因此,在数据的采集和加工方面,要考虑到数据来源的选择(期刊文献数据库或书籍、网站信息、新闻报道等);要制定科学的采集策略和采集方式(手工或自动);数据的规范化处理;数据的临床特征和临床证据特征的标引的可行性和进行方式(自动或人工)等。要充分考虑到以上各步骤实现的可行性、难易程度和时耗,据此进行必要的取舍,设计专门的数据加工工具,提高加工效率。

3.3 系统的功能设计要充分考虑到系统的定位和不同的用户需求

系统的定位不同,系统的设计就会有很大差别。因此,要特别注意系统的定位,特别是在系统检索功能的设计上,要特别注意突出主要需求,并从不同用户的角度出发设计出他们最乐意使用的功能模式^[4,5]。比如,对于临床医生而言,他们更乐意从临床学科划分或证据类型等专业参数入手进行检索;对于患者而言,更多的则是从疾病名称或治疗方法等非专业参数入手进行检索。

因此,系统设计中要特别注意:各种不同检索模式的检索功能的划分;一些常用的重要功能的在不同模式下的必要重复和一些不常用功能取舍;各种模式下检索入口参数、检索条件和检索策略的设计;检索结果的延伸以及检索结果的输出模式等的设计等,都要根据该模式的使用者需求习惯仔细推敲确定^[4-5]。

3.4 数据库的设计要充分考虑到系统功能和效率双重因素

循证医学数据库必须为用户提供检索平台,供临床医学工作者查找其感兴趣的数据。那么数据库系统的设计不仅必须足以支撑系统的不同检索方式、检索功能和其他系统功能的实现,还必须要支持系统的高效率运转(如检索效率、计算速度、使用方便程度等)。例如,在库结构的设计上,可以考虑将多种不同类型的证据集中在一个数据表中,也可以分类分别建立数据表。前者的数据库结构简单,但数据表有一定的冗余,在数据量特别在时会检索效率产生影响;后者则突出了证据个性特征,便于用户使用、系统效率也会较高,但数据库表结构关系复杂,对证据数据的采集和标引加工提出了更高的要求。又

如,不仅要给不同用户提供不同的检索模式和检索入口、限定条件等参数,还要提供保证查全或保证查准等对用户透明的工作机制(后台检索策略的修正);基于诸如疾病名称、药品名称等同义词表的词义扩展等智能化检索模式,使用户只需要进行简单的界面参数选择,就能实现复杂的检索过程。

3.5 数据库系统的设计要充分考虑到作为产品的诸多实际问题

作为一个研究课题,系统的研究可以只为建立系统模型,不考虑系统数据的更新、持续化的数据生产等实际问题。但作为一个要投放市场的完整的正式的产品,仅进行一次性的数据搜集加工整理或仅建立系统模型是没有实际意义的,必须要考虑到数据库中数据的持续生产、加工、发布和更新;要考虑到用户合法性认证、产品使用模式、管理方式、产品的版本控制和宣传销售模式;要考虑到系统和数据的安全性、知识产权保护措施等一系列实际问题。例如,循证医学数据库产品可以采取光盘发布的模式,但这种模式下则无法实现对用户的个性化信息增值服务,如新文献提醒服务等。在互联网版的产品中可以方便地实现个性化服务功能,但产品中必须增加用户帐户管理、访问控制模块(IP控制或其他限制模式)、不同类型资源的使用权限和计费方式等。又如,无论是哪个版本的产品,必须对系统和数据进行加密,保护自己的劳动成果不被非法盗用。

4 研发工作体会

随着循证医学知识的普及和循证医学实践活动的开展,循证医学数据库作为一种新型的重要的信息产品,已经被很多以生产销售信息产品为主业的有实力的公司和研究机构所看中,也受到越来越多用户的青睐。循证医学数据库的学术价值和市场前景不言而喻。

中文循证医学数据库的建设是一个量力而行、逐步完善的过程。数据库的早期版本可在系统模型的建立和用户需求的探讨上多下功夫,此后可根据实际情况逐步扩大科学证据的收录范围,改善加工工具,完善相关的辅助信息服务系统,开发更多的增值信息服务功能,逐步建立完善的专业及非专业用户的循证医学资源获取与利用、学术探讨与信息交流的综合平台。还可以充分利用Web2.0技术和信息服务模式,在系统中引入用户反馈机制,使循证医学证据检索系统具有智能化“自主学习”功能和较强的用户交互能力,使循证医学数据库产品更为成熟,更适应市场需求,形成用户与产品之间相互作用的良性循环,不仅为循证医学证据的获取和充分利用创造了更好的条件,也将在最大程度上为循证医学实践活动提供方便。CM

参考文献

- [1] 刘传和.循证医学实践——医学图书馆面临的机遇和挑战[J].中华医学图书馆情报杂志,2005,14(4):9-11.
- [2] 王秀平,贺培凤.网络循证医学信息的检索与应用[J].中华医学图书馆情报杂志,2005,14(6):53-55.
- [3] 郭庆文.如何获取循证医学证据[J].医学信息,2005,18(9):1098-1101.
- [4] Wiley InterScience:Reference Work;the Cochrane library 2007,issue 4.[2007-11-18].<http://www.thecochranelibrary.com>.
- [5] PubMed Clinical Queries [EB/OL].[2007-11-18].<http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml>.