图书情报工作>>
第 56 卷 第 9 期 2012 年 5 月

基于数据的中医药知识服务研究*

高 博 崔 蒙** 杨 硕 贾李蓉 董 燕 朱 玲
中国中医科学院中医药信息研究所 北京 100700

〔摘要〕论述“馆所合一”的中医药专业图书馆发展知识服务的必要性,阐释现阶段进行基于数据的中医药知识服务的基础和下一步发展所必需的相关建设,指出数据知识服务是比数据服务级别更高的数据利用手段,也是知识创新、知识发现的有力支撑,主要可分为定制知识服务和扩展知识服务,而知识服务平台则是同时容纳二者的开放性用户界面。目前,真正意义上基于数据的中医药知识服务尚未形成,下一阶段的重心应是人才培养和多学科团队建设。

〔关键词〕知识服务 数据 中医药
〔分类号〕G350.7

Knowledge Services of TCM Based on Data

Gao Bo Cui Meng** Yang Shuo Jia Lirong Dong Yan Zhu Ling
Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700

〔**Abstract**〕The paper illustrates the necessity for a professional library of Traditional Chinese Medicine(TCM) to develop knowledge services, and expounds the foundation of TCM knowledge services based on data at the present stage and the necessary relevant construction at the next step. Data knowledge services are the method that is better than data services and the strong support for knowledge innovation and knowledge discovery knowledge services include customized knowledge services and expansion knowledge services. And knowledge services platform has simultaneously held the two's open user interface. At present, knowledge services of TCM based on data are not formed, and the personnel training and the multi-disciplinary teams building are the cores on next stage.

〔**Keywords**〕knowledge service data Traditional Chinese Medicine

知识服务是以信息知识的搜寻、组织、分析、重组为基础,根据用户的具体问题和个性化环境,直接融入用户解决问题的过程,能够提供有效支持知识应用和知识创新的服务[1]。联合国开发计划署(The United Nations Development Programme, UNDP)将其简练归纳为“基于全球先进知识上的建议、专长、经验和试验方法等,帮助咨询用户获得解决问题的最佳方案”[2]。20世纪90年代末知识服务的概念被引入国内,引发了国内对知识服务的研究热潮,但迄今为止,国内尚未形成真正意义上成熟的知识服务。

中医药图书馆属于专业图书馆,建立在研究院所的中医药图书馆,也就是通常所说的信息所与图书馆“馆所合一”的图书情报机构,主要服务于科研机构的高素质研究人员及已具备相当专业基础的研究生。他们从网上获取信息资源的能力非常强,特别在本专业领域内,获取电子资源的能力甚至不亚于专业图书馆人员,因而很少会到馆使用阅览室或电子阅览室。因此,中医药“馆所合一”的专业图书馆,其知识服务主要体现在知识平台的构建与服务以及相关标准研制等工作上,而基于数据的知识服务则是我们目前主要的研究领域。

1 数据知识服务

1.1 数据知识服务的基础

知识服务是根据用户的具体问题,对数据进行检索、筛选、清洗、处理,最终提供给用户其所需结果的过程,因而在知识服务的发展初期,数据的收集和处理是最基础的手段。针对这一需求,中国中医科学院中医药信息研究所

* 本文系国家高技术研究发展计划(“863计划”)项目“中国中医药科学数据网格服务应用”(项目编号:2006AA01A122)研究成果之一。

** 崔蒙系本文通讯作者。

收稿日期:2012-03-14 修回日期:2012-04-10 本文起止页码:5-9 本文责任编辑:王善军

联合全国37家中医药及相关学科高等院校、研究院所,已经建成国内规模最大的中医药专业数据库群[3],涉及中医、中药、针灸、古籍、民族医药等各个领域,包括中药化学实验数据库、中药药理实验数据库、临床疾病数据库、临床个案数据库、针灸临床医案数据库等[4]。其特色是构建了基于文献拆分的结构型数据库群,使文献数据的挖掘成为可能,为基于数据的知识服务奠定了坚实的基础。目前这些数据库已经经过了整合、规范和在互联网上共享[5],并提供大量的关联检索结果。

由此可见,基于数据的中医药知识服务,主要资源是文献数据,主要模式是结构型数据库,现阶段的研究重心应该转移到如何利用结构型文献数据进行知识服务上来。

1.2 数据服务

数据服务是最初级的数据利用手段,即将已有的数据直接提供给用户,由用户根据自身需要,对数据进行有目的的检索和筛选,从中选取自己需要的数据资料,然后根据研究目的对数据进行再加工,并最终获取自己所需的知识。这也可以看作是知识服务的雏形。

这一方法要求用户本身有极高的专业素养,同时具备一定的情报学和信息学基础。首先,用户需要明确自己的需求,并对研究目标进行精准定位;其次需要用户懂得数据清洗与数据拆分,使加工后的数据符合自身需求;最后还需要用户具有数据处理能力,以便知识发现与知识创新。该服务模式对数据的利用率极低,用户必须对数据进行全面检索,才能达到自己的目的;而其他用户即使需求相同也必须重复这一系列步骤。这种数据服务模式,对用户要求高、步骤繁杂、数据利用率低,只是知识服务的一种初级探索。

1.3 数据知识服务

数据知识服务是较高级别的数据利用手段,即由数据拥有者根据用户需求对大量数据进行有目的的筛选,再将结果提供给用户的服务模式。这一模式又可分为有方向性的定制知识服务和无方向性的扩展知识服务。

1.3.1 定制知识服务

定制知识服务是根据用户需求,以用户欲解决的问题为目标,不仅为用户检索并提供数据,更要根据相关知识对提供的数据进行筛选、清洗、拆分、重组,构建相应的结构型数据库,提供适当的算法与工具,提出解决问题的方案。

这一服务模式非常重视用户需求分析,比起提供用户需要的数据,它更加关注通过服务解决用户实际问题。因而,定制知识服务不仅要充分解读用户提出的问题,更重要的工作是协助用户构建恰当的问题。在进行研究设计时提出的问题是否恰当,关系着研究是否具有重要意义,并决定着解决问题的方向和解决方案的制定[6]。选题过于宽泛,则研究方向不够明确,研究结果没有针对性;选题过于细化,则可获取数据资源太少,造成研究结果偏差、缺乏应用推广价值。因此,作为知识服务者,应该参与到用户提出问题、寻求答案、解决问题的全过程中去。

以中药新药研发为例:笔者所在研究所(以下简称“我们”)在为中药新药研发单位提供基于数据的定制知识服务时,反复与用户讨论,明确其需求;然后根据需求,设计数据库结构,选取适当的数据处理方法及工具,基于中药单味药、中药药理、中药化学、中药方剂、中医疾病等数据库,筛选数据,进行数据清洗,建立具有西医疾病、西医病理、中医证候、中医方剂、中药饮片、中药药理、中药化学成分等数据元的结构型数据库,形成数据挖掘平台,获得初步结果[7];再与用户及相关专家讨论,根据专家意见进行数据及工具的调整,最终获得用户所需的新药处方,实现知识发现与知识创新。

1.3.2 扩展知识服务

扩展知识服务针对无具体问题,以学习知识、拓展知识面为目的的用户,针对用户意欲拓展的知识领域提供较为科学的研究方向和相关数据资料。

这一服务模式需要对用户领域知识结构有一定了解,结合用户意欲拓展的领域方向、深入程度,根据领域现有的知识结构、专业基础、学科框架和发展需求等方面进行扩展知识服务。该服务主要向用户推荐与其研究主题相关的知识元素,依据领域知识网络,拓展用户的知识视野,以满足用户学习新知识、拓展知识面的需求。相对于定制知识服务而言,扩展知识服务对用户需求分析的要求略宽松,对知识服务者的知识深度要求下降,但对知识广度的要求更为严格。

我们研制的中医药学语言系统(Traditional Chinese Medicine Language System,TCMLS)能够很好地发挥扩展知识服务的作用。TCMLS是参考美国国立医学图书馆开发的统一医学语言系统(Unified Medical Language System,UMLS)研制的,它由语义类型和语义关系组成语义网络,能够建立起所有概念间的逻辑关联关系,形成中医药学的概念体系甚或知识体系,与UMLS共同应用,能够在很大程度上满足用户对扩展知识服务的需求。虽然其所形成的概念或知识体系仅仅包含了概念间的线性关系,更多的非线性关系未能

包括在其中，但这对用户的扩展知识服务已经能够起到很好的支撑作用：首先TCMLS解决了同一概念的同义词与近义词问题，使用户用任意术语检索均可获得其所表达的概念以及表达该概念的所有术语所关联的知识；其次当用户需要时，通过语义关系可以获得该术语所表达概念的所有相关概念，而这些相关概念通常是人脑很难考虑周全的。

目前TCMLS已收集了30万个词汇和11万个概念，通过108种语义类型和68种语义关系，建立起超过127万个的语义关联关系。这个初步建立的语言系统已应用于中医个案数据库并关联了PUBMED和百度检索，成功地为领域或非领域用户提供了扩展知识服务。

1.4 人员分析

在现阶段要快捷有效地满足用户需求，由专业人员基于数据、针对用户提供知识服务，是比将已有海量数据直接提供给用户更为先进的手段。一来专业人员可以快速判断出所需数据和筛选、清洗原则，二来面对相近的需求和问题可以进行合并处理，大大减少了无效数据处理。这些对知识服务提供者的知识结构和知识层次有很高要求。首先，知识服务提供者应具备良好的情报学素养，能够迅速定位适当的数据资源、数据结构、处理工具。其次，知识服务提供者必须具备多学科知识，对服务面向的学科领域有较深造诣，明确学科定位及知识结构，并对相关专业和交叉学科有一定深度的涉猎。

但知识服务面向的受众很多，分布于各个专业；受众基础不一，有些正在入门阶段，有些已经在进行最前沿的研究；他们对知识服务提供者的服务深度和广度也要求不一。要求每个服务者同时达到所有条件，会对服务者造成太大压力。因而知识服务不应局限于一对一，最高效的模式是团队合作，在用户提出需求时，根据需求所属学科、涉及学科，由团队中相应专业人才进行问题设计及解答。这样一来，服务人员虽然也需要对多学科有所涉猎，但术业有专攻，无需全部精通，从而可以缩短服务人员的前期培养时间，并且减轻服务人员的负担。

我们在开展和提供基于数据的知识服务时，基本依靠团队。这种团队是建立在具有虚拟机构、协同工作、汇集共享基础上的虚拟研究团队，即虚拟研究院。采用这种模式，是因为中医药图书情报机构的编制都偏小，独立承担行业知识服务研究和提供的能力均显不足；另外中医药图书情报机构本身的人员组成结构也很难适应基于数据的知识服务的需要，其信息处理与人工智能的知识都显不足。上述原因导致中医药领域开展基于数据的知识服务研究很难在有围墙的独立研究机构内完成，而需要由基于网格技术的多学科、多研究单位组成虚拟研究机构来完成。这种虚拟研究团队的长期、稳定建设保证了中医药基于数据的知识服务的研究与提供。

2 知识服务平台建设

知识服务平台是团队型知识服务的体现，是由知识服务团队建设的，能够实现多对多知识服务的开放性用户界面。

2.1 数据绑定型知识获取平台

学科特点决定了中医药数据属于知识密集型数据。占主导地位的传统中医药数据依然是文献数据，包括古代文献与现代文献，因此其研究目标不是建立数据密集型科研模式，而是直接面向基于数据的知识发现与知识创新。

数据绑定型知识获取平台可以帮助用户从数据中更有效地获取知识，该平台上部署了大量结构型数据库以及用于这些数据库的数据处理工具。这些数据库一般是基于现代或古代文献数据加工后的、针对某一明确目的的、具有良好关联关系的结构型数据库，根据用户需求对数据进行了清洗，完成了同义词与近义词的处理以及古今词汇的转换，改进和完善了数据处理工具。数据经过处理后，可以产生用户所需的新知识，供用户在进行中药或临床研究时参考。

2.2 知识服务平台

知识服务平台基于数据绑定型知识获取平台建设而高于数据平台，是以知识服务为目的的开放性用户界面，因而必须以满足知识服务者的多学科知识储备为前提。多数据库联合成数据库群，并随时可以纳入新的数据库是知识服务平台的首要前提，我们已经部署在数据绑定平台上的大量结构型数据库基本可以满足这一要求。

第二个必须满足的条件是智能搜索，其中又包含了模糊检索、检索词推荐、扩展性检索等。用户的情报学基础参差不齐，因而要令服务平台具备模糊检索能力，并给出合理的检索词、检索式推荐，将用户检索结果稳定在一个合理范围内；此外还需进行扩展性检索，提供相关学科专业的关联性研究，以便开拓思路、创新知识。我们已经进行的数据清洗，对同义词、近义词的处理以

及古今词义的转换等先期工作对此构成了良好的支撑。

第三大功能是建立模拟应答系统，收集用户需求及问题构建问题库，由专业人员给出解答。该应答系统可部分代替人工服务，对用户的初级问题进行解答，避免了知识服务者对低级问题的反复回答。用户提问的过程同时也是应答系统收集问题的过程，通过解答用户问题，不断充实问题库，可以令应答系统越来越高效地满足用户需求。

第四，定制服务依旧是知识服务的重中之重。中医药学发展到现在，分科越来越细，学科的突破性进展越来越难，交叉学科的发展也越来越受到重视，前沿的研究则普遍涉及了多学科内容，需求复杂。对于这一部分用户必须启用团队服务，由用户提出需求，定制服务，然后由知识服务团队进行综合解答，解决问题。

知识服务平台的第五个特点是可嵌入。除了访问服务平台的网站之外，在其他系统内都可以嵌入部分服务，并随时可以链接入网站。如临床嵌入式知识服务，即在电子病历页面里嵌入知识服务窗口，给出相似度较高的病历及相关文献记载等，供临床医生诊断、处方时参考。当遇到疑难情况需要更多知识支撑的时候，可以直接从嵌入式窗口跳转入服务平台，进行扩展性搜索。

知识平台建设成功后，可以应用于教学、科研，尤其是探索性、发现性科学研究，有效地支撑知识创新和知识发现。目前，中医药知识服务平台的构建已经初具规模，我们将其称为“中国中医药科学数据网格”，其基本构想是在数据网格平台上绑定大量数据及数据处理工具，留有用户接口。用户既可以使用平台上的数据及其工具进行数据处理获取知识，也可以调用平台数据使用自己的处理工具处理数据、获取知识，还可以将自己的数据传输到平台上使用平台工具进行处理以获取知识。该平台主要包括一个体系、两个系统和两个应用平台。

中医药语义本体系统是广泛计算架构和领域知识集合的结合，整个构架分为两个层次：核心网格服务层和虚拟语义视图层。前者主要指网格中的资源以及直接基于自主开发的“中医药数据网格”（DartGrid）服务实现的服务，后者指的是用于支持上层本体应用的虚拟语义视图。该系统向用户提供了一个基于浏览器的统一浏览查询接口，而本体的分布式结构对终端用户来说则是完全透明的，呈现在用户面前的是一个虚拟的大规模中医药学领域本体。目前，由本体网格支持的中医药学语言系统已经在线发布，实现了上层语义网络的发布和浏览，用于支持上层面向中医药的知识应用，包括了数十万个概念和近百万条实例，基本上覆盖了整个中医药领域的十几个子学科。

面向中医药领域的分析算法与知识集成的系统是根据数据的特性与不同用户的需求，在中医药数据挖掘、知识发现算法与服务的基础上，研究和建立中医药数据知识发现管理子系统“中医药数据网格挖掘系统”（DartSpora），从海量的数据中识别出有效的、新颖的、潜在有用的知识。DartSpora 继承了开源数据挖掘软件在数据挖掘方案控制方面的先进思想和丰富算法资源，并结合以网格计算和云计算为代表的并行计算技术，实现了该数据挖掘服务平台，并以平台的挖掘服务为基础开发了多种面向中医药领域的数据挖掘应用。目前，DartSpora 平台包含了 6 个中医药数据定制算法和 209 个常用数据挖掘算法，形成了一个数据挖掘算法资源库，并在计算集群上以计算服务的方式对外提供算法服务，方便科研人员。

基于中医药领域语义本体的文献标引与加工平台是一个文献协作平台。该平台利用笔者所在研究所研究的中医药文献标引规则和《中国中医药学主题词表》，按照不同的划分标准，对文献进行不同角度的标引，大大提高了数据加工的速度和效率，便于文献的统一集成及分析。

中医药搜索平台是我们面向动态中医药专题自动生成所研发的，在全文搜索引擎的排序机制上考虑关系数据表的特性，实现了各类数据库与数据平台的一体化检索。同时开发了基于矢量标记语言的网络矢量图绘制工具包，对相关的搜索网站进行聚合和分析，提供图片链接。

中医药科学数据网格服务体系的建立能够改变中医药研究人员对信息的取用方式，能改变信息的整合方式，从而改变查新、统计、研究分析的效率、规模，为中医药规模性研究、知识积淀和应用推广奠定基础。同时，针对中医药信息网格所制定的一系列的标准和规范，能够进一步完善基于中医药信息网格的数据共享与利用。

3 基于数据的中医药知识服务发展机遇

知识服务作为知识经济时代一种新的知识应用理念，针对具体而实际的需要解答的问题，提供全面有效的信息增值服务。这是一种带有前瞻性的研究活动，相对数据服务、文献服务、信息服务等而言，在资源收

藏、信息处理、服务方式等方面都发生了质的超越[8]。这种超越性大致体现在四个方面:一是个性化,即用户可以根据自己的实际需求选择知识的范围和层次;二是知识性,即以内部的知识共享机制为基础,依靠团队化合作,为用户提供相关的知识单元内容;三是创新集成化,即对专业学科资源和最新进展进行创新性的提炼整合后,将资源集成在一个界面,然后提供给用户;四是专业化,即以提供专业化知识为中心,整合相关专业知识,提供知识服务。

由这些特点来看,真正的知识服务特别是基于数据的中医药知识服务,目前还远未实现,但发展知识服务已经迫在眉睫。不仅因为数据资源积累到一定程度后需要高质量的应用模式产生,单就中医药学自身发展而言,其对知识服务的需求也已经达到了一个临界点。

中医药学是一门历史悠久的学科,经过几千年发展已经形成了极其庞大的学科体系;新学科、新科技出现后,与中医药学不断融合,又形成了大量新兴交叉学科。围绕这个庞大体系产生了巨量的文献、知识,独特的个体诊疗和灵活的辨证论治又令知识点产生了海量的排列组合方式,以个人的力量,终其一生也不可能完全掌握。以临床诊断而言,为了找到最贴合疾病本质的证候,临床医生需要从病人全部症状体征中寻找出最本质的病机,同时不能放弃兼加证候,并根据主证、次证的轻重程度调整处方。在这一过程中,如果能找到文献支持是最佳选择,但医生单凭个人的经验和知识储备很难从汗牛充栋的中医药学文献中提取出对自己有用的信息。而知识服务,特别是嵌入式知识服务可以根据相似度比较,提供相近的病例文献,启发医生的思路。

在知识服务的发展过程中,有两个关键点:①前期的关键是知识储备,即原始资料积累,这一步已经通过数据库的大量建设等工作基本达到要求,后期需要时也可以随时扩充;②现阶段的关键则是知识整合,不仅是检索、收集,而且要将结果整合成一个体系,满足用户需求,解决用户问题。因而,知识服务者必须进行思考,对检索、收集来的数据进行分析、提炼、重组,改变数据服务时代单纯"拿来主义"式的应用思路。在这一过程中,少不了多学科的团队协作。

由此可见,基于数据的中医药知识服务现阶段的短板是缺乏专门进行知识服务的人才(尤其是中医药学以外的专科人才),缺乏一支多学科知识服务人才组建的团队。因而,发展中医药知识服务,下一阶段的重心应该转移到人才培养和多学科团队建设上来。

参考文献:

[1] 张晓林. 走向知识服务:寻找新世纪图书情报工作的生长点[J]. 中国图书馆学报,2000(5):32-37.

[2] What are UNDP's "knowledge services"[EB/OL]. [2010-11-25]. http://www.undp.org/execbrd/pdf/UNDP%20knowledge%20services.pdf.

[3] 崔蒙,谢琪. 中医药数字化研究进展. 医学信息学杂志,2008(10):13-15.

[4] 刘静. 建立中医药数据服务与利用平台[J]. 世界科学技术-中医药现代化,2009,11(4):582-584.

[5] 崔蒙,尹爱宁,范为宇,等. 中医药科学数据建设研究进展[J]. 中国中医药信息杂志,2006,13(11):104-105.

[6] 顾骏. 基于循证科学理念的知识服务模式[J]. 理论与探索,2011,34(10):22-23.

[7] 雷蕾,张慧敏,崔蒙,等. 中医药化学辅助研发系统的建设. 中国中医药信息杂志,2008,15(8):100-101.

[8] 刘旭东. RSS技术在数字图书馆知识服务中的应用[J]. 情报科学,2011,29(11):1684-1687.

〔作者简介〕高 博,女,1981年生,助理研究员,发表论文10篇;崔 蒙,男,1953年生,研究员,博士生导师,发表论文100余篇;杨 硕,女,1975年生,副研究员,发表论文10余篇;贾李蓉,女,1977年生,助理研究员,发表论文10篇;董 燕,女,1976年生,研究实习员,发表论文11余篇;朱 玲,女,1979年生,助理研究员,发表论文10余篇。