

DOI: 10.3969/j.issn.1671-3982.2012.01.016

• 现代图书情报技术 •

医学知识共享平台的构建

刘 明, 李 焱, 王洪军

[摘要] 了解放军医学图书馆数据库研究与开发部自主开发的医学知识共享平台, 详细描述了医学知识共享平台的组成和架构、知识服务的设计与实现、Web 服务的实现与部署, 提出了图书馆自主开发知识服务数据库的思考。

[关键词] 生物医学; 知识共享平台; 数据库; 知识服务

[中图分类号] R318; G250.74

[文献标志码] A

[文章编号] 1671-3982(2012)01-0047-04

Construction of medical knowledge sharing platform

LIU Ming, LI Yan, WANG Hong-jun

(Medical Library of Chinese PLA, Beijing 100039, China)

[Abstract] After an introduction of the medical knowledge sharing platform developed by Department of Database Research and Development, Medical Library of Chinese PLA, and a detail description of its composition, architecture, design and implementation of knowledge and Web service, some ideas on the independent development of knowledge service database in library were proposed.

[Key words] Biomedicine; Knowledge sharing platform; Database; Knowledge service

知识共享平台^[1]是知识信息资源和数字化学习与研究的中心, 是文献检索、网络学习、数字化研究中具有全新沟通机制与丰富资源的学习与研究环境。它借助电子商务的经营管理方法, 为文献检索、数字化学习、数字化研究和网络教育服务, 是现代信息技术、网络技术和远程教育发展的需求。

医学知识共享平台是解放军医学图书馆数据库研究与开发部集中解放军医学图书馆已有资源和部门部分产品, 通过集中整合而自主开发的包含医学图书、期刊、知识库、科学数据集和应用工具的综合服务平台。系统分两个阶段实施, 第一阶段重点是文献库和知识库^[2]的建设, 第二阶段重点是科学数据集等新型数据库的研制以及应用工具的开发。

1 医学知识共享平台的组成和架构

医学知识共享平台集中整合了解放军医学图书馆数据库研究与开发部的多项开发成果, 统一移植和构建于 SQL Server、Lucene^[3] 和 .NET 开发平台之上。医学知识共享平台的组成和架构见图 1。

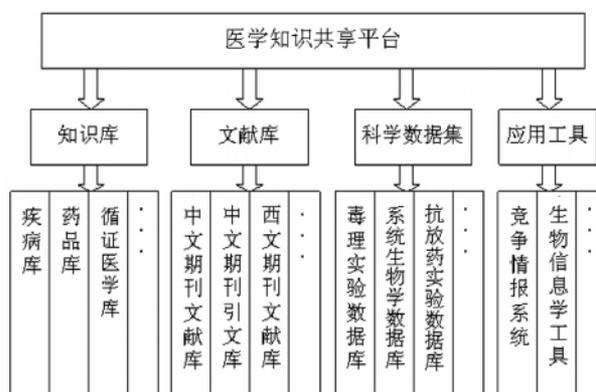


图 1 医学知识共享平台的组成和架构

[基金项目] 国家自然科学基金面上项目(81001699)

[作者单位] 解放军医学图书馆, 北京 100039

[作者简介] 刘 明(1977-) 男, 满族, 辽宁沈阳人, 博士, 副研究员, 主持国家面上课题 2 项, 参编著作 1 部, 发表论文 20 余篇。

1.1 知识库

医学知识库是根据各学科、专业的知识结构将医学图书进行分类整合, 以疾病为知识核心, 扩展到

药品、辅助检查、循证医学证据等相关知识而形成的网络知识服务平台。它将数字化过程与发掘和增值书刊的知识资本结合起来,变革和创新书刊出版模式,促进信息资源管理向知识管理发展。它将拓宽学习医药学知识的途径,使学习变得更加便利。

医学知识库主要包括疾病数据库、药品数据库、手术学数据库、辅助检查数据库、循证医学证据数据库、疾病研究新进展数据库、手术图谱数据库、临床操作规范数据库、医保药品数据库以及其他知识产品与服务等,通过知识点之间的内在联系将各库有机地结合起来,建立各库之间的知识链接^[4-6]。

1.2 文献库

文献库是以解放军医学图书馆已开发完成的 CMCC(中文生物医学期刊文摘数据库)、CMAC(中文生物医学会议数据库)、CMCI(中文生物医学引文数据库)等文献数据库为基础,为用户提供各类生物医药中西文期刊文献的服务系统^[7-8]。

CMCC 是解放军医学图书馆研制开发的中文生物医学文献书目型数据库,也是目前检索国内生物医学文献最常用的数据库之一。自 1994 年创建以来,在中国内地和港澳拥有广泛的用户,是国家和军队卫生部门认可的重要检索工具之一。收录中文期刊库累积期刊文献 470 万篇,年递增 40 余万篇,半月更新。学科范围包括医药、卫生、生物以及与之相关的文化教育、军事医学、环境、化学、农业、自然科学等。

1.3 科学数据集

数据库以及数据库系统在药物设计领域中应用广泛,不但可以作为存储、检索、管理化合物的平台,而且还能作为研究分子结构、活性的重要工具。科学数据集主要收集国内外研究机构的实验数据以及文献发表的相关实验数据,实现基于药物类别、名称、结构和子结构的检索,对减少科学家重复实验、加速药物的开发具有重要的意义^[9-10]。

1.4 应用工具集

应用工具部分主要包括基于科学数据集之上的生物信息学工具以及分子结构工具,基于文献之上竞争情报分析和自助式的定题服务等。

2 知识服务的设计与实现

2.1 知识元是构建知识库的基本元素

知识是由不同的知识单元组成的,知识单元包含若干片段知识,片段知识是由多个知识元构成的。一个个单独的知识元依据其逻辑关系组成知识单元,知识元之间所依存的逻辑关系称为知识元链接。一种知识体系是由不同的知识单元组成的,每个知识单元又是由不同的独立知识元通过知识元链接排列组合而成的。不同知识之间可以通过知识元链接形成不同的知识链,进而编织成整个知识结构的知识网络。知识库凭借知识元链接在知识网络中实现对任意知识元和知识单元的存取(见图 2)。

2.2 文献知识元的抽取是知识仓库构建的基础

若将每篇文献看成是人类知识结构中的一个知识单元,知识单元与知识结构就组成了个性知识与共性知识的知识系统。个性知识体现在知识的创新上,共性知识结构则表现在知识的完整性上。

2.3 知识元标引是对知识仓库进行知识管理的基础

文献知识的标引与检索以前只停留在以整篇文献为单元,选取若干主题词,或采用引文索引,或只是采用文本知识发现提供文献中精细的信息和多篇关联的信息。知识仓库的知识标引通过对文献的知识元进行标引,揭示文献所包含的知识元与相关学科知识结构的链接关系,从中找出信息与知识导航链接的普遍规律,为用户直接获取有效知识元提供最有效的方法。知识元标引起着知识元过滤和知识元链接的作用。

医学知识体可分为若干个知识单元,如疾病知识、药品知识、辅助检查知识、循证医学证据、手术学知识等。每个知识单元中包含有大量的知识元信息,如疾病知识单元中包含各种各样的疾病,药品知识单元中包含多种药品,辅助检查知识单元中又包括相当数量的检查项目。

通过对 Lucene 源代码的修改,将我们设计的算法嵌入其中,使其具有医学专业词表自动生成与维护、自动下载与分类网络上循证医学文献的功能,成为更适合于中西生物医学的检索软件。并以此为基本开发平台,构筑涵盖疾病、药品、实验室检查、循证医学的临床医学应用软件。

知识库技术路线见图 2。

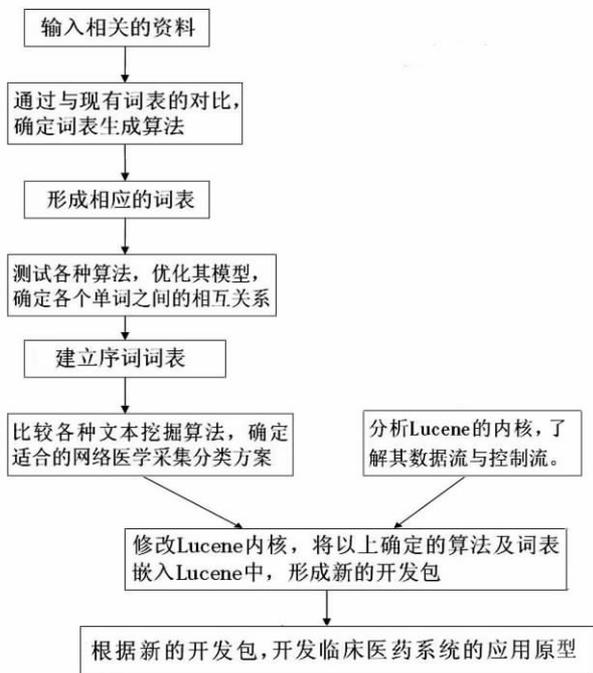


图 2 知识库技术路线图

3 Web 服务的实现与部署

系统的主要功能有用户服务系统和后台管理系统两大部分。

3.1 用户服务系统

用户服务系统包括用户注册与登录,用户信息查询与维护,疾病知识库内容检索、浏览和下载,中西文期刊文摘检索与浏览,全文文献申请,我的文件夹的查询及维护,用户卡充值管理(余额查询及充值),个性化定制服务,论坛,综合信息检索与浏览等(见图 3、图 4)。



图 3 知识库内容检索、浏览和下载



图 4 中西文期刊文摘检索与浏览

3.2 后台管理系统

管理服务系统包括用户基本信息管理(用户信息查询与修改)、信息流量统计分析、全文传递系统维护基本信息维护(文献上传、文献发送、计费)、充值卡销售与管理(售价、批发)、系统留言信息管理、系统信息统计(用户类别与地区分布统计分析、文献类别统计分析、期刊分类统计及分析、数据库使用情况统计、出错情况统计、用户资金流向统计、系统费用年/季/月报表)等后台管理系统(图 5)。



图 5 后台管理系统

4 讨论

医学知识共享平台有四大优势:一是整合了解放军医学图书馆的多项信息资源而建设的数字化学习与研究中心;二是以知识元抽取和知识元标引为基础的知识服务实现;三是依据生物医学领域研

研究的发展建立了以科学数据为基础的新型数据库; 四是以文献库和科学数据集为基础, 在数据之上开发的应用工具。系统分两期实施, 第一期重点是文献库和知识库的建设, 第二期重点是科学数据集等新型数据库的研制以及应用工具的开发, 目前第二期正在进行中。

作为医学图书馆, 在建设数字化图书馆和自主开发知识服务的过程中, 一要立足于本图书馆的传统和优势, 突出重点, 形成自己的优势; 二要根据国内外医学信息相关数据库的现状、部门自身的发展历史和现状、人才队伍的结构等, 选择能在国内外产生影响或研究方向独特的研究领域形成特色数据库; 三要在医学和信息学的发展呈现出不断分化和综合的趋势, 新兴学科、交叉学科和综合学科不断涌现的情况下, 使医学信息服务不断跟踪学科发展的前沿, 寻找新的学科增长点, 开发更多新的数据库产品。

【参考文献】

- [1] 孙卫, 罗之兰, 张蔚. 科学数据共享平台的数据管理研究[J]. 科学与科学技术管理, 2005, 26(12): 32-36.
- [2] 李娜, 张文举. 浅谈知识化医学图书馆[J]. 中华医学图书情

报杂志 2006, 15(3): 7-9.

- [3] 管建和, 甘剑峰. 基于 Lucene 全文检索引擎的应用研究与实现[J]. 计算机工程与设计, 2007, 28(2): 489-491.
- [4] 张玉, 张文举, 李娜. 构建以知识服务和知识组织为基础的医药学知识库[J]. 医学信息学杂志 2010, 31(2): 26-29.
- [5] 张文举, 李娜. 基于知识服务的医学知识服务系统研究[J]. 中华医学图书情报杂志 2007, 16(5): 1-5.
- [6] 李娜, 张文举. 医药学知识咨询系统构建过程中相关概念的探讨[J]. 中华医学图书情报杂志 2006, 15(6): 67-69.
- [7] 张晓梅, 杜云祥, 雷春炳, 等. 基于网络的生物医学信息导航系统的研发[J]. 中华医学图书情报杂志 2006, 15(6): 12-15.
- [8] 雷春炳, 张晓梅, 颜世刚, 等. 西文生物医学期刊文献数据库 (EMCC) 建设[J]. 现代图书情报技术 2005, 21(8): 54-57.
- [9] Tyrkk E, Pelander A, Ojanper I. Differentiation of Structural Isomers in a Target Drug Database by LC/Q-TOFMS Using fragmentation Prediction [J]. Drug Test Anal (S1942-7611), 2010, 2(6): 259-270.
- [10] Riner A, Chan-Tack KM, Murray JS. Original Research: Intravenous Ribavirin - review of the FDA's Emergency Investigational New Drug Database (1997-2008) and literature Review [J]. Postgrad Med (S0032-5481), 2009, 121(3): 139-146.

[收稿日期: 2011-08-26]

[本文编辑: 杜云祥]

(上接第 43 页)

其次, 域模型的构建完全基于已知的知识信息, 虽然在一定程度上实现自定义表单可覆盖病历信息, 但在临床中遇到复杂病历时, 其内容难以完全用预先定义的域模型进行描述, 只能添加自由文本类型的节点进行描述。这也与 OpenSDE 的初衷有所差异。

【参考文献】

- [1] 丁宝芬. 医学信息学[M]. 南京: 东南大学出版社 2009.
- [2] 伟斌, 肖强, 陈联忠, 等. 电子病历系统的研究与开发[J]. 中华医院管理杂志 2004, 20(4): 204-206.
- [3] Shortliffe EH. Biomedical Informatics [M]. New York: Springer 2006.
- [4] Schnipper JL, Linder JA, Palchuk MB. "Smart forms" in an Electronic Medical Record: Documentation-based clinical decision Support to Improve disease management[J]. J Am Med Inf Assoc (S1067-5027) 2008, 15(4): 513-523.
- [5] 雷健波. 电子病历的核心价值与临床决策支持[J]. 中国数字

医学 2008, 3(3): 26-30.

- [6] 于一, 廖睿, 叶大田. 电子病历结构化方法概述[J]. 北京生物医学工程 2007, 26(1): 103-106.
- [7] Johnson S B, Bakken S, Daniel D *et al.* An electronic Health Record Based on Structured Narrative [J]. J Am Med Inf Assoc (S1067-5027) 2008, 15(1): 54-64.
- [8] Los R K, van Ginneken A M, van der Lei J. OpenSDE: A Strategy for Expressive and Flexible Structured Data Entry [J]. Int J Med Inf (S1386-5056) 2005, 74(6): 481-490.
- [9] Los R K, van Ginneken A M, Marcel de Wilde. OpenSDE: Row Modeling Applied to Generic Structured Data Entry [J]. J Am Med Inf Assoc (S1067-5027) 2004, 11(2): 162-165.
- [10] Los R K, van Ginneken A M, van der Lei J. Extracting data recorded with OpenSDE: Possibilities and Limitations [J]. Int J Med Inf (S1386-5056) 2005, 74(6): 473-480.
- [11] 肖强, 吴伟斌, 陈联忠, 等. 电子病历系统的主要实现技术[J]. 中华医院管理杂志 2004, 20(4): 207-209.

[收稿日期: 2011-08-09]

[本文编辑: 王天津]