

元数据技术在医学信息资源整合中的应用

尚 慧 李雪琴

(山西医科大学图书馆 太原 030001)

〔摘要〕 对生物医学信息资源及元数据进行分析,并在此基础上深入剖析元数据技术在不同医学信息资源整合中的应用,包括 OPAC, Web Service, SFX, OAI, 知识本体系统等,提出利用元数据技术整合医学信息资源时应注意的几个问题。

〔关键词〕 信息资源;整合;元数据

Applying of Metadata Technology in Medical Information Resource Integration SHANG Hui, LI Xue-qin, Library of Shanxi Medical University, Taiyuan 030001, China

〔Abstract〕 The paper analyzes biomedical information resources and metadata. Based on that, it goes deep into analyzing the metadata technology application in different medical information resource integrations, including OPAC, Web Services, SFX, OAI, knowledge ontology system. At last it proposes several problems which should be noted while applying metadata technology to integrate medical information resources.

〔Keywords〕 Information resources; Integration; Metadata

随着网络技术和数字化技术的发展,医学信息资源日益丰富,各医学文献部门除拥有自建的数据库外,大多购买数十个以上电子数据库(电子期刊、电子图书、网络数据库等)。由于这些数据库检索平台和检索语言各不相同,有的甚至存在较大的差异性,不同的检索界面和检索方式给读者检索和获取信息带来诸多不便。通过元数据互操作技术将异构的资源有机地整合在一起,可以有效、系统地满足人们对医学信息的需求。

1 生物医学信息资源及其整合

1.1 生物医学信息资源

生物医学信息资源包括医学图书、期刊、专利文献、会议资料、学位论文、临床实验报告、病历报告、手术操作规程、临床照片、影像资料等等,它们可能是纸本形式,但越来越多的是数字形态,且往往以多种媒介形式存在。主要包括在线医学数据库、医学搜索引擎、网站医学资源和医学专业电子期刊,通过对这些资源的分析,发现生物医学信息资源具有与一般信息资源同样电子化、网络化、动态性、分布性、无序性的特点^[1]。

1.2 生物医学信息资源的整合

信息资源的整合,就是将数据库资源为中心的各种文献资源通过中间技术处理,整合一体,统一在一个平台上,提供一站式服务,以提高资源利用效率^[2]。医学信息资源整合能消除信息孤岛,优化

〔收稿日期〕 2010-06-10

〔作者简介〕 尚慧,硕士,馆员,发表论文9篇;李雪琴,硕士,副研究馆员,发表论文10篇。

医学数字资源环境,提升知识关联功能,满足用户个性知识需求,医学信息资源的整合程度直接关系到医学信息资源能否被高效利用。当前,国内医学信息资源整合还处于起步阶段,主要集中在大学图书馆和专业的出版社和数据库公司,比如同方,维普等。这些整合还只是处在数据库集成阶段,大多在各个数据库之间建立链接,没有将各个数据库中的数据做到语义整合,没有形成真正自己的整合数据库。用户在使用过程中常常会遇到断链或查找不到资源等问题,而且不同数据库中相同的资源会造成数据的冗余,不利于资源的有效利用。

2 元数据互操作技术

2.1 元数据

元数据其实就是“描述数据的数据”,或者“关于数据的结构化数据”^[3]。元数据是用来描述数据本身的内容特征和其它特征的数据,是医学数字信息组织和处理的基本工具,它们为各种形态的医学数字化信息单元和资源集合提供了规范、普遍的描述方法,元数据整合中开放描述和互操作技术已成为一个基本要求。

2.2 元数据的互操作技术

2.2.1 概述 经过几十年的发展,在网络数据资源的交换与共享领域,元数据已经从最初对信息对象的描述和抽象扩展成包括数据发现、数据转换、数据管理和数据使用的整个网络信息组织利用过程中必不可少的工具和方法之一,借助于元数据技术可以实现异构系统之间的互操作。元数据互操作是指多个不同元数据格式的释读、转换和由多个元数据格式描述的数字化信息资源体系之间的透明检索^[4]。

2.2.2 元数据映射 两个元数据格式间元素直接转换,或通过中介格式进行转换。

2.2.3 标准描述方法 其过程是建立一个标准的资源描述框架,用这个框架来描述所有的元数据格式。

2.2.4 元数据复用 通过在一个元数据格式中,

引用其它元数据格式的部分元素或属性,用来描述复杂资源,从而扩展元数据格式适用范围,以促进元数据的相互转换。

2.2.5 元数据开放搜寻 利用元数据来进行资源搜寻和发现时,可以采取元数据开放搜寻机制来实现元数据互操作,典型机制是 OAI。

2.2.6 元数据语义转换 通过元数据语义定义和元数据概念集,支持两个元数据格式间元素通过语义分析进行转换。

2.3 元数据互操作技术类型

2.3.1 基于核心元数据的互操作 在医学元数据方案中根据各自描述的内容对象特征,目前普遍选择 DC 作为核心内容格式,选择其它元数据格式作为扩展内容格式,通过元数据复用可以快速建立元数据格式和辅助元数据转换^[5]。

2.3.2 基于元数据标准映射的互操作 映射是元数据基于内容互操作最基本的方式,有动态和静态两种方式。动态方式采用元数据转换中间件,将相应的查询请求中的有关内容转换成资源站点支持的元数据模型或可以识别的元数据格式,返回时再转换成本系统支持的元数据形式,以支持本系统的查询结果处理。静态转换类似于目前的搜索引擎,将资源站点的数字对象尽可能多地抓取到本地,按照本地的元数据模型建立索引,提供服务。

2.3.3 基于资源描述框架的互操作 资源描述框架(Resource Description Framework, RDF)是一个能对结构化元数据进行编码,交换及再利用的体系框架。RDF 定义了一个由资源、属性和陈述 3 种对象组成的基本模型,人们可以利用 RDF 来解读所引用的元数据,使不同的用户或团体在这一框架下定义他们自己的元数据元素,从而提供各种不同的元数据体系之间的互操作性。

2.3.4 基于协议层的互操作 实现元数据互操作技术的协议有 Z39.50 协议、OpenURL 协议、OAI-PMH 协议等。这些协议使得计算机之间以一种标准的、相互可理解的方式进行通讯,并支持不同数据结构、内容、格式的系统之间的数据传输,实现异构平台异构系统之间的互联与查询。

3 元数据技术在不同信息资源整合中的应用

3.1 OPAC 系统中的元数据技术

在网络环境下, MARC 格式通过 856 字段链接到医学电子资源的实际网页地址 URL, 将它整合到 OPAC 系统。这样读者在利用 OPAC 进行医学信息检索时就可同时检索出印刷型和数字型资料, 并链接到电子资源的网址, 实现在线浏览。这种技术实现起来比较简单, 但局限性很大。MARC 注重形式、结构复杂、制作成本高的局限, 使其难以承担对大量多媒体数字资源进行编目的重任, 也无法完全实现知识资源的立体化和网状化整合。而且这种技术是静态链接, 无法实现机构端本地化控制, 更新和维护比较困难, 链接对象的任何细微变化都容易形成死链。

3.2 Web Service 中的元数据技术

Web Service 是一种信息发布和访问技术, 它是一个采用 XML 通过 URL 来发布接口和应用绑定的软件体系结构。其主要特点有: (1) 良好的封装性、标准性及开放性。(2) 高度的可集成性, 完全屏蔽了软件平台的差异, 实现了当前环境下最好的互操作性和集成性。(3) 利用 XML, 便于扩充, 具有很强的描述能力^[6]。因此, 它非常适于异地异构数据源集成, 用其构建的虚拟的全局数据库, 能有效地整合医学信息资源。

3.3 SFX 资源整合中的元数据技术

SFX 是 Special Effects Cineatographlly 的缩写, SFX 实际上是一个基于开放的统一资源定位器标准的上下文相关的参考链接系统^[7]。SFX 作为一种比较成熟的整合技术, 通过 OpenURL 框架把复杂的数据库之间的互连通过简单的链接完成, 实现检索数据“一步到位式”参考链接, 从而在异构的分布式数据库系统之间实现无障碍导航, 理论上使用户获得所有可获得的资源。其 OpenURL 也已经成为美国国家标准信息组织 (NISO) 标准。

3.4 OAI 整合中的元数据技术

这是一种完美的资源整合形式, 能够实现对网上所有异构的医学数据库资源进行检索, 它要求所有的资源拥有者都遵循一定的协议, 只要资源在网上, 用户就一定能够检索到。标准资源整合成本最低, 服务最有效。开放文档先导 (The Open Archives Initiative, OAI) 推出的元数据收割协议 (The Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH) 提供了一个元数据互操作框架。它的目标是发展和促进互操作标准, 促进内容数据的有效传播。它提供给用户的是统一元数据格式的数据。从用户的角度看, OAI 服务提供者所起的作用是对一定类型的信息资源进行整合并提供查询服务。

3.5 知识本体 (Ontology) 元数据技术

知识本体是近几年来语义万维网领域研究热点, 元数据可以提供医学资源属性的描述集合, 每个资源对象可以基于不同目的、不同角度, 采用不同的医学元数据方案进行描述。这些不同的元数据标准或方案都是基于医学领域的某种认识和看法的概念体系 (规范词表和类表等)。这种明确的、形式化、规范化和可共享的概念体系就是知识本体^[8]。因此有学者认为本体就是元数据之元数据。知识本体既然是明确的、形式化和规范化的, 因而可以方便地建立不同知识本体间的映射、参照机制, 从而实现不同的医学元数据标准以及医学资源对象间的普遍联系, 实现不同元数据标准或方案间的语义互操作, 实现医学领域知识的共享和重用, 大大拓展人类处理和利用医学知识的能力。

4 元数据技术在医学信息资源整合中应注意的问题

4.1 制定元数据标准

在当前医学信息资源建设过程中并没有执行统一的元数据标准, 元数据的形式多样。由于元数据标准不一, 兼容性差, 医学信息资源整合困难。所

以要跟踪元数据发展,选择国际上通用的、比较成熟的元数据方案,积极参与制定元数据标准,完善对医学数据资源的描述和各种元数据格式之间的兼容和互操作研究。

4.2 积极借鉴国外经验

元数据整合国外要比国内起步早,可以借鉴国外经验,根据本国情况加强对元数据与医学数字对象有机整合的途径与方法研究、元数据技术规范以及相关技术体系研究,建立医学信息资源元数据体系,构建以元数据体系为中心的医学资源整合的管理体系。

4.3 逐步构建元数据系统与概念体系

通过元数据技术可以实现对数据层的整合,但随着被整合网站在数量和信息量方面的增加,大量数据会对网络医学资源的同步搜索、发现、更新、分析处理效率产生影响,因而要逐步构建医学信息资源元数据系统,来规范元数据,促进元数据管理,减少重复生产。推进研究利用元数据进行基于知识的医学数据组织和知识发现,建立统一、标准

的概念体系,为实现医学信息系统语义互操作打下扎实基础。

参考文献

- 1 韩夏,吴廷照. 生物医学信息资源的元数据描述与整合[J]. 情报学报, 2006, 25 (3): 312-315.
- 2 马文峰. 数字资源整合研究[J]. 中国图书馆学报, 2002, (4): 64-67.
- 3 张婷. 浅析元数据在特色数据库建设中的应用[J]. 情报杂志, 2006, (1): 136-137.
- 4 李纪宾,邹小农,乔友林. 元数据及其在医学科学数据共享中的应用[J]. 医学情报工作, 2006, (1): 10-14.
- 5 王卷乐,游松财,孙九林. 地学数据共享中元数据扩展和与操作技术[J]. 兰州大学学报(自然科学版), 2006, 42 (5): 22-24.
- 6 何蕾. WEB 信息资源整合技术及其系统实现[J]. 计算机应用, 2003, (6): 59.
- 7 李富玲,卢振波. SFX-信息资源整合新工具[J]. 现代图书情报技术, 2002, (6): 69-71.
- 8 贾宏. 数字图书馆网络信息资源整合及其技术分析[J]. 情报杂志, 2006, (2): 39.

(上接第 34 页)

6 结语

互联网的魅力很大程度在于它能源源不断地提供最及时的信息。信息是网络之所以存在和发展的基础。网站的生存和发展也必须遵循这一规律。网站的生命力不在于创建,而在于维护更新。不断策划新活动、新主题是网站的重要生命力之所在。人们上医院网站就是要获取所需要的医疗信息。医院网站只有不断地提供人们真正需要的医疗服务和保健需求的内容才能有吸引力和生命力。所以,一个以制度建设为保障而不断更新、稳定运行的医院网站才会有长远的发展,才会给医院带来真正的社会效益、经济效益。

参考文献

- 1 梁珂. 论医院网站的定位与设计[J]. 中国医院管理, 2004, 24 (10): 45-46.
- 2 杜文霞,朱东屏,陈贞春. 军队医院网站建设与应用[J]. 解放军医院管理杂志, 2006, 13 (7): 607-608.
- 3 谷敏. 医院网站的建设与管理[J]. 医学信息学, 2005, 18 (2): 107-108.
- 4 网络营销知识. 网站维护的基本内容及基本知识[EB/OL]. [2010-01-05] <http://www.seovea.com/wang-zhanyouhua/195.html>.
- 5 于丽艳,朱有禄,孙玲. 新形势下医院网站建设与发展探析[J]. 医学信息学, 2009, 30 (12): 61-63, 85.
- 6 蔡宏伟,高东怀,曹小曼. 大型医院网站技术设计[J]. 中华医院管理杂志, 2003, 19 (8): 508-510.
- 7 王振东,魏冰玉,房靖祥,等. 医院网站设计思路及前景展望[J]. 中华医院管理, 2007, 27 (2): 63-64.
- 8 薛梅. 如何创建一个成功的医院网站[J]. 现代医院, 2004, 4 (4): 95-96.