

图书馆元数据管理及应用浅谈

——以湖北省图书馆为例

涂寓 段超

湖北省图书馆 湖北 武汉 430060

[摘要]以湖北省图书馆为例,由湖北省图书馆数字资源现状及存在的问题,谈到通过元数据仓储的搭建管理元数据。在湖北省图书馆元数据仓储的基础上加强了数字资源元数据管理,实现了数字资源统一检索,提高了数据服务质量。

[关键词]数字资源;元数据;元数据仓储;统一检索

【DOI】10.12252/j.issn.2096-6261.2020.02.411

1 湖北省数字资源现状

随着互联网和数字图书馆在中国的迅速发展,数字资源在图书馆的馆藏及服务的作用越发重要。多年来,湖北省图书馆对数字资源投入了大量经费,通过自建数据库及购买外购数据库等手段,积累了相当一批数据库。湖北省图书馆利用不同的资源加工软件,建设了长江讲坛资源库、湖北戏剧志、三国文化、红色历史文化资源库、辛亥革命专题资源库、非物质文化遗产资源库等四十余个专题数据资源库;同时购买商业数据库如CNKI数据库、万方知识数据库、维普期刊数据库等五十余个。

数字资源库的种类和资源量迅速膨胀,使得读者在享受海量数据带来的丰富知识的同时,也深受种类繁多和数量庞大的数字资源的困扰。如何将海量数据安全的保存、有序的归类、高性能的检索提取、成为新的问题。为了从整体上规划我馆的资源建设,避免重复浪费,早在我馆数字图书馆建设之初,资源元数据仓储的建设就纳入了规划。

2 湖北省图书馆元数据仓储平台的搭建

元数据仓储平台是实现图书馆资源元数据管理与应用的有效工具。它可以完成抽取及管理各类基础资源的元数据信息的任务。

我馆元数据仓储平台支持基于模板的资源元数据库扩展与自定义,满足我馆各类异构信息资源的整合与组织,以及对信息管理与信息服务,知识管理与知识服务的应用需求。

通过建设本馆自有的元数据仓储平台,可以有效的管理本馆资源元数据,使得检索更高效;可以用于支持资源扩展利用,为信息深度利用处理服务提供基础;可以为专题库构建提供元数据基础,方便快捷、灵活的组织专题文献。

2.1 系统架构

元数据仓储平台根据各资源元数据规范生成元数据模板,通过配置好的采集适配器从不同来源、不同结构的数据源中抽取元数据信息,实现元数据的采集、存储、管理和发布。

元数据仓储平台的功能应包括元数据采集、元数据存储、元数据管理、元数据发布、元数据接口。图书馆需要制定规范的元数据标准以及搭建功能完备的元数据仓储,通过搭建元数据仓储平台对元数据进行清洗和管理,不断聚集优质的元数据,才有可能建成灵活实用的知识资源服务体系,进而对读者进行知识服务而不是简单信息展示。

2.2 元数据采集对象分析

能够被提取元数据信息的数据库都能成为元数据采集的对象。包括:自建数据库、分布式数据库以及商业数据库。

2.2.1 自建数据库

本馆自建数据库的资源一般是图书馆内部资料或征集来的具有对外服务版权的一类资源,湖北省图书馆通过现有的著录平台进行加工整理而形成的专题或特色数据资源库,因此这类资源库的数据结构以及数据库的相关信息可以直接掌握,较易获取元数据。

对于不同结构的数据库,可以分别按照其元数据必备字段以特定格式要求导出元数据,为了易于统一管理,我们通常导出xml文件。

2.2.2 分布式自建数据库

分布式自建数据库指数据资源以及对象文件不在本地,一般为其他合作机构建立的数据库。

对于这类数据库可以采集其元数据信息保存到元数据仓储,与本地元数据整合,建立统一索引库。当读者用户通过元数据仓储检索到想要的资源信息,并访问其详细内容时会向本地WEB服务器发送访问请求,通过统一的索引库,如果访问的对象数据在本地,则直接将本地数据返回到发布页面。如访问对象不在本地,则直接将异地数据库中的数据提取到本地发布端。由于使用了统一的索引库,并不需要把异地数据中的数据取存放在本地存储服务器中,节省了存储空间,提高了访问效率。

2.2.3 商业数据库

我馆购买了大量商业数据库,这类数据库通常由不同的数据库商维护管理,我们很难获得其元数据信息。

经过多年的深入合作与积极洽谈,有越来越多的商业数据库资源商愿意提供部分元数据及相关视图或检索接口,我们可以根据电子资源供应商提供的开放服务方式,通过直接获取xml文件,或者Web Service接口、相关API接口等获取该电子资源供应商授权的元数据信息。

2.3 元数据采集

在采集元数据前,需要根据各类资源的元数据建立对应的元数据模板,设定相应的元数据字段名称及数据类型。这项工作是通过数据网关来实现的。

元数据获取首先要制定统一的元数据规范并严格遵守。

《元数据规范》对元数据仓储的表结构及字段名称进行规定,使得各个异构源数据库能够统一字段管理,便于数据同步。

2.4 元数据管理

元数据管理主要包括3个部分:元数据录入、元数据抓取、元数据同步。

2.4.1 元数据录入

根据预定义的元数据结构,由管理员对不同资源的元数据进行录入。其功能要求应包括:

系统具有灵活、可定制的元数据加工模板,根据定制的元数据模板实现对元数据的著录。;

对每条元数据记录系统自动生成元数据唯一标识(MetaID);

系统能根据需要自动生成相关字段,如记录唯一标识;能在元数据模板中设置缺省值,如语种;

对具有数字对象的记录,系统能与相应的数字对象关联,且能与对象数据处理系统结合自动生成部分元数据并提供人工干预的功能;

对具有数字对象的记录,系统能为每一个数字对象自动生成数字对象唯一标识符;

一条元数据可以关联多个数字对象数据;

支持对中文MARC和西文MARC的批量导入功能。并可对导入的记录添加字段以及与其相关的全文或附件;

支持对其他数据库文件的批量导入功能。并可对导入的记录添加字段以及与其相关的全文或附件;

支持多种分类的自动互相映射,一般有中国图书馆分类法、学科分类、专用分类表,用户利用系统设置映射类,实现多种分类自动生成;

在数据加工和处理中,能够在多窗口中互相切换。

2.4.2元数据抓取

元数据仓储平台应针对不同的数据来源提供相应的数据网关和采集适配器来实现对元数据的采集和处理。

资源采集模块能够根据系统预定义的元数据结构获取目标资源库资源的元数据信息。通过数据网关提供的接口,各个采集适配器及数据库之间进行数据交换。采集适配器可以根据数据源提供的接口定制,负责采集对象数据库中的元数据,并生成新的数据结构存储到元数据仓储中。

2.4.3元数据同步更新

在我馆元数据仓储建设早期,元数据采取非实时、被动的导入,需要人工定期上传更新元数据,不仅耗费大量的时间,数据更新时效也非常滞后。同时我馆自建专题数据库建设时间跨度非常大,建库所用的技术手段不统一,因此结构并不一致,如果想要数据库更新时,自动上传数据到元数据仓储服务器,需要将每个数据库按照统一标准修改,为此我们做了大量的基础工作。通过查看数据库的表关系,在源服务器上根据表的关系导出数据,接着根据表的关系在本地机器上删除数据,最后在本地机器上通过命令导入数据。

目前符合我馆元数据标准的数据库,除了可以手动导入,还可以通过接口同步的方式将元数据自动同步到元数据仓储内。元数据库服务器会定时查找对象数据库是否有更新,如果有更新,则进行增量同步。

2.5元数据库存储

元数据仓储需要采集并存储大量电子资源与外部网络资源的元数据信息,因此考虑到数据存取效率和存储安全,元数据仓储架构以分布式进行部署,存储服务器可以分配在不同的服务器中,由此,可以方便实现数据的扩展,提高并发用户的访问性能以及检索服务的可靠性,利于实现本地管理。

2.6元数据管理

数据管理是元数据仓储的核心功能,因此元数据库仓储应具有良好的维护性,对于数据的变化,例如,数据库增加新内容或者增加新的数据库,元数据仓储管理端都应该能够应对自如。

2.7元数据发布

元数据发布应支持灵活定义元数据发布模版。在仓储平台系统中预先配置各类资源的通用发布模板,在不同的模板中配制允许检索和发布的元数据字段。通过管理模块,可配置元数据更新、策略、发布的策略以及元数据模板管理等。同时,通过配置相应的检索功能,实现对不同元数据项的信息检索和组合检索,能够多维度分类查看元数据仓储中不同元数据库的相关资源,发现不同元数据之间的关联。

2.8元数据接口

元数据仓储系统需要提供标准API,支持应用集成和二次开发,以保障系统的可扩展性以及数据的可复用性。

3元数据应用场景

通过元数据仓储平台的搭建对图书馆各系统中的不同数据进行统一分类,有效整合,针对不同业务系统软件及数据的特点,制定不同的数据采集方案,将不同系统的业务数据推送至数据仓储,进行分类集中存储、分析运算以及数据应用。形成全馆数据资源的汇集中心,继而通过API接口的方式向其他数据应用系统提供数据基础。

针对本馆数字资产进行统计,对各类资产汇总显示,可以得知文献资产数据和运行日志等情况。

通过采集并清洗本馆馆藏数据、部分提供元数据的数据库商以及自建数据库的元数据,实现本馆自建资源以及纸电的统一检索;按年度或指定时间范围生成图书馆各类服务报告,形成可视化的展现,如汇总图书馆年度资源建设情况、文献资源总量情况、图书借还情况、数字资源利用情况、读者入馆情况、门户访问情况、读者办证情况、咨询服务情况等。

4结束语

湖北省图书馆的对元数据的管理及应用只是一些初步的探索与尝试。面对海量数字资源以及读者对知识服务日益增长的需求,图书馆将不可避免地遇到利用数字资源优化馆藏并提供高质量的信息服务的问题。

通过持续的采集和分析数字资源服务数据,在有数据仓储平台数据支撑和相关平台支撑的基础上,进一步挖掘分析,形成数字资源的推送服务。帮助图书馆利用数据找到存在的问题,协助图书馆通过数据做好读者精准服务,更加高效利用现有的数据资源,利用先进的信息技术服务读者,是我们永远孜孜以求的目标。

参考文献

- [1]张宏伟,齐明明,史磊,李杨,基于元数据仓储的统一检索系统研究,图书馆学刊,2014,第9期;
- [2]陈明,图书馆员专业化建设对信息服务的推动作用,图书馆学刊,2014,第6期;
- [3]韩永进,汪东波,孙一钢,魏大威,李东来等,中国图书馆事业发展报告 数字图书馆卷,北京,国家图书馆出版社,2017