

可嵌入用户环境的 IR 框架设计¹

吴登禄^{1,2} 祝忠明²

(1 中国科学院研究生院 北京 100080; 2 中国科学院国家科学图书馆兰州分馆 兰州 730000)

【摘要】 通过对机构库以功能为导向框架的分析,对目前的以功能为导向的机构库进行机构,形成了以服务为导向的机构库框架,并提出了机构库的服务嵌入到用户环境的实现方法,以此来说明,机构库可以为用户提供主动式的服务,方便与其他系统的无缝集成和互操作。

【关键词】 基于服务的机构库框架 嵌入式服务 以用户为中心 Web2.0

Design of the architecture of Embedded Institutional Repository for User Environment

Wu Denglu^{1,2} Zhu Zhongming²

(1 Graduate University of Chinese Academy of Sciences, Beijing 100080, China; 2 The Lanzhou Branch of the National Science Library of the Chinese Academy of Science, Lanzhou 730000, China)

【Abstract】 This paper analyzes the function of the institutional repository (IR) based on function and deconstructs it. Then, the authors described the framework based on service and supply the methods which realize the service of IR embedded in the user's environment. On this basis, it is shown in this paper that IR can supply the initiative service with user and be easier integration and interoperability with other system.

【Keywords】 the service architecture of the institutional repository Embedded service User-centered Web2.0

1 引言

随着互联网技术的发展, Ajax 技术, Blog, RSS 等的出现,使互联网从 Web1.0 时代跨入了 Web2.0 时代。在 Web2.0 的环境下,信息环境变得越来越开放^[1],对于数字图书馆的服务组织,将超越以图书馆为中心的服务组织模式,而向以用户为中心的服务组织模式转变,数字图书馆的各种服务及数字图书馆应用系统的设计和开发,必须考虑如何方便地被集成到、嵌入到用户的信息空间和工作流程中去。

机构库^[2] (Institutional Repository, 以下简称 IR), 作为一种重要的数字资产保存系统, 也应该考虑使其嵌入到用户的工作环境和信息空间中, 实现数字化知识资产的统一管理, 长期积累和保存, 来集中展示和反应科研机构的研究实力和水平, 拓展学术成果的发布和交流渠道, 提高学术成果被发现和引用的几率, 扩大研究人员及研究所的学术影响和声望。

随着数字图书馆服务模式的变化和用户环境的变化, 对机构库的功能和服务

¹本文系中国科学院国家科学图书馆“全院联合机构仓储体系建设”项目及国家社科基金项目“机构知识库建设研究与应用”(项目编号: 07BTQ019)的研究成果之一

作者简介: 吴登禄(1983-),男,甘肃张掖人,硕士研究生,从事计算机信息检索和处理、数字图书馆技术与系统研究; 祝忠明(1968-),男,甘肃永昌人,研究员,硕士生导师,从事网络信息系统组织建设、数字图书馆技术与系统的研究开发工作。

也提出了挑战。目前 IR 的架构都是一种功能性的架构, 以提供功能为主, 用户只能登录到机构库中才能获得相应的服务, 与用户的工作环境相互割裂, 分割了与其他系统的交互。使机构库以一种孤立的形态在发展。

基于上述背景, 本论文通过对机构库以功能为导向框架的分析和研究, 对传统的以功能为导向的 IR 进行解构, 形成了以服务为导向的 IR 框架。以服务为导向的 IR 设计, 可以支持将 IR 的服务嵌入到用户的工作环境中去, 实现以 IR 库为中心到用户为中心的服务组织模式的转变。

2 功能导向的 IR 框架

功能导向的机构库, 作为一种资产保存系统, 主要是保存数据和对数据的读、写、删除操作以及数据的检索, 因此从功能上说具有以下核心功能, 如果缺少其中的一种的功能就不能充分的支持机构库。

2.1 数字资源的提交和收集功能

机构库作为一个机构的学术成果的数字存储, 其中的数字资源来源于机构人员的提交和从外界获得, 因此必须通过数字资源的提交和收集功能来增加系统资源。用户可以分为普通用户和管理员, 普通用户可以通过 Web 方式提交数据, 提交的内容包括数字对象的元数据信息和内容信息, 管理员可以通过批量进行数据的导入。

2.2 数字资源的存储和保存功能。

机构库获得数字资源之后, 为了实现数字资源的检索和利用, 机构库一般有资源的存储和保存功能。支持各种类型资源的保存, 包括文本、图像、音频和视频等, 并提供数据资源的备份和恢复。

2.3 数字资源的浏览和检索功能

建立在资源保存功能之上, 用户通过浏览和检索功能来发现机构库中的资源。就现有的 IR 而言, 用户通过 Web 方式来实现资源的检索, 检索方式有作者的姓名, 资源名称、发布日期、类型、关键字和全文检索等。

2.4 互操作功能。

互操作功能是机构库的一项重要功能。随着用户环境的变化, 各种专业的学术网站和各种门户的出现, 信息的搜索却在不同的环境中, 因此降低了用户的检索效率。实现系统之间的互交换, 可以整合不同系统之间的资源, 便于用户检索和利用, 节省了时间提高了检索的效率, 并且增加了系统的资源。例如采用 CNRI 的 handle^[3] 系统来唯一标识数字对象, 支持多种元数据格式, 如 DC、MODS 和 METS 等。支持 OAI 协议^[4], 实现元数据的暴露和收割。

2.5 管理功能

管理功能提供对机构库系统的管理, 包括用户权限管理(例如管理员和普通用户、多种认证方式)、资源管理、资源使用统计、基于工作流的资源审核和发布管理。

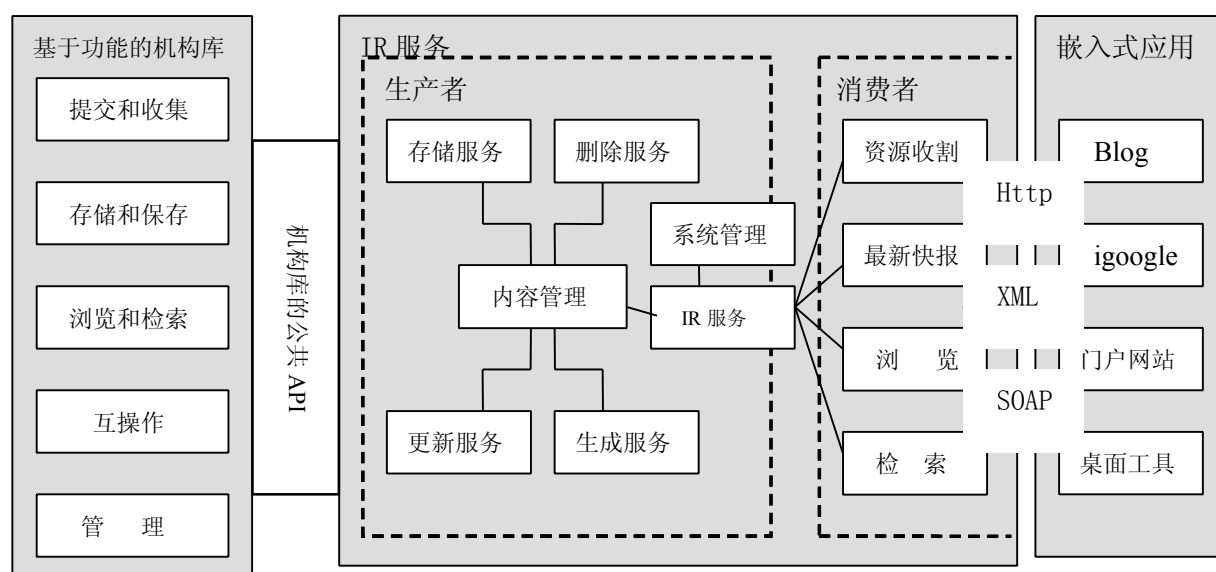
用户信息环境和信息获取行为的变化, 促使图书馆系统的开发, 应该强调以用户为中心, 资源的组织强调用户主导, 与用户的互动性强调主动性, 即使不是用户创造资源, 也由用户附加资源。功能导向的机构库, 实现了数据的保存和检索等功能, 但它是一种功能性的架构, 模块之间具有很强的耦合性, 模块不能单独构成应用, 与其它系统的集成性很差, 功能不能实时嵌入用户的环境, 因此对

其功能导向的框架进行解构，形成以服务为导向的架构，在其基础上实现 IR 功能的可嵌入，使 IR 的各种功能以服务的方式融入用户的信息空间和环境。

3 可嵌入用户环境的机构库框架设计

目前的 IR 是一种基于功能的系统，实现了包括数字资源的提交和收集、数字资源的存储和保存、数字资源的浏览和检索、互操作、管理五种功能。虽然目前以功能为导向的 IR，在网络化和信息化的环境下为方便学术资源的存取和促进学术交流方面发挥了积极的作用，但随着用户信息环境的不断发展和系统之间互相集成趋势的不断增强，以功能为导向的 IR 逐渐不能适应这种发展趋势。可嵌入用户环境的 IR 建立在基于功能的机构库框架之上，框架设计的主要思路：首先对目前以功能为导向的机构库进行解构，在目前机构库的基础上，对其功能性的 API，以 Web Service^[5] 的方式进行封装，形成一种面向服务的架构^[6] (service-oriented architecture, 以下简称 SOA)，实现基于服务的机构库；在基于服务的机构库基础上利用 Javascript 和 Gadget/Widget^[6] 技术来实现机构库功能的可嵌入，使其融入用户的信息环境和工作流程中去。

可嵌入用户环境的 IR 由 3 部分组成：基于功能的机构库系统、基于服务的机构库系统、嵌入式应用。如图 1 所示：



1 基于服务的机构库框架

3.1 基于功能的机构库框架

包括数字资源的提交和收集、数字资源的存储和保存、数字资源的浏览和检索、互操作、管理功能^[7]，作为机构库的基本功能，用户要必须登录到其环境中才能得到其服务，它们都提供公共的 API 接口。

3.2 基于服务的机构库框架

在 Web2.0 时代，信息资源将是泛在的，用户需要的信息资源也将是泛在的，用户希望不离开自己的科研和生活环境就可以随时随地的获取所需要的资源和服务，因此对机构库所提供的服务也应该是实时的和可嵌入的，提供面向用户环境的嵌入式服务。基于服务的机构库是建立在基于功能的机构库框架之上的，即利用基于功能的机构库的公共应用程序接口来实现基于服务的机构库的功能，形

成一种面向服务的架构。

SOA 致力于在因特网或企业内联网上将应用程序功能作为服务进行提供，它淡化了不同系统之间所形成的界限。机构库可以将所具有的核心功能作为 Web Service 来发布，以这些 Web Service 来构成应用，形成机构库的服务框架来实现机构库功能的实时性和可嵌入。

如果把 IR 看成一种资源，那么对其进行的操作可以用读、写、删除、更新来概括。IR 的服务从功能上可以从两个角度进行划分：资源生产者和消费者。向 IR 提交资源的用户称为资源的生产者，从 IR 中获取资源的团体和个人称为资源的消费者。

从资源生产者的角度来看，机构库的服务主要有：

(1) 内容管理服务

机构库作为一种存储系统，其中的资源是机构库的核心内容。资源的生产者向机构库提交资源，机构库接受资源并进行存储和实施管理，是机构库的核心功能。基于功能的机构库系统，资源的生产者必须登录到其环境中才能对其实施资源的提交和管理，使得用户脱离了自己的工作环境，大部分时间是在不同的系统之间进行切换，浪费了大量的时间降低了工作的效率，基于服务的机构库通过对其解构，把基于功能的机构库的内容管理通过 Web Service 的方式提供服务，形成了内容管理的实时性，资源生产者可以不用离开自己的工作环境随时随地的提交资源。内容管理服务包括：

① 资源生成和存储服务 (Creation Service)：资源的生成和存储通常是联系在一起。资源要进入机构库，必须满足机构库的存储格式。在基于功能的机构库环境中，资源的生成是通过资源提交流程来完成的，比如首先经过身份验证，再填写许多关于提交资源的元数据信息等，生成满足机构库存储格式的数据之后才能提交资源。在基于功能的机构库中，上述过程是在机构库的环境中完成的，提交流程要经过好多过程才能提交成功，提交流程的复杂，所需时间长，最主要是生产者必须登录到其环境，使得生产者必须离开自己的工作环境，不方便资源生产者的使用。对基于功能的机构库解构之后，其资源的提交已经被包装成了一种 Web 服务，生产者不用离开自己的工作环境，就可以提交信息包，比如可以把资源的生成和存储服务做成一个桌面小工具，嵌入到生产者的工作环境当中，随时的提交信息包。资源的生产者通过身份验证后，利用桌面工具来生成机构库资源所需格式的信息包，然后通过机构库的存储服务，提交信息包。

② 资源更新和删除服务：资源生产者如果要对其提交的资源进行修改，在基于服务的机构库中，可以脱离机构库的环境，在自己的工作的环境当中进行资源的更新和删除。

(2) 系统管理服务

系统的管理服务，是为了方便资源的生产者对系统实施管理的，比如登录信息，一个研究社团资源的管理和组织。

从资源消费者的角度看，机构库的服务有：

① 资源收割服务 (Obtain Service)：机构库系统一方面通过资源的生产者来提交资源来增加机构库的资源，另一方面可以通过获取其他系统的资源来增加系统的资源。同时也暴露资源被其他的系统所收割，资源的收割可以增加系统之间的互交换，扩大资源的共享范围和增加机构的影响力。国内外各种实用的机构库几乎都遵循 OAI-PMH 协议，以 OAI 数据提供者的身份开放自己的资源，提供资源的收割服务。

② 最新快报服务(Alerting Service): 最新快报可以提供给一个机构最新提交的资源, 资源消费者如果关心一个机构的发展动态和某一研究领域的最新研究进展, 可以通过机构库提供的最新快报服务来实现, 在基于功能的机构库中, 资源的消费者必须登录到其环境中才能了解该机构的发展动态和某一研究进展, 因此为了方便用户, 把最新快报封装成一种快报服务来提供给资源的消费者, 例如以 RSS 的方式来提供非消费者。请求一个 RSS 的资源, 返回 XML 格式数据, 因此可以很方便的和其他系统集成。

③ 浏览服务(Browsing Service): 资源的生产者提交资源, 机构库存储资源。浏览功能是机构库的一个重要的功能, 使资源被很好的发现和利用, 比如按照作者、提交的时间、关键字、标题等浏览。浏览功能是资源消费者在利用机构库是使用的比较多的功能。因此, 把基于功能的机构库的浏览功能包装成 Web 服务, 对其功能进行封装, 利用 Javascript 技术对其浏览功能进行包装或用 Gadget/Widget 技术以一种桌面工具的方式来嵌入用户的环境, 为用户提供实时服务和很好的方便系统集成。

④ 检索服务(Search/Retrieve Service): 检索是机构库中发现资源的最重要的手段。资源的消费者要查找其中的资源, 一个快速的方法就是使用检索功能, 系统之间的集成也需要检索功能的支持。但目前以功能为导向的机构的检索功能, 从资源消费者的角度来看, 消费者必须进入其环境才能检索, 从系统集成的角度来看, 与其他系统的集成变得很困难。而在 Web2.0 的环境下信息都是泛在的, 因此资源消费者的环境是一种泛在的环境。可以利用 SRU^[8]来实现其检索服务, SRU 是一种 Web 服务的检索协议, 并在其基础上实现检索服务的嵌入。因此机构库的检索功能作为一种 Web 服务来实现包装, 可以适应资源消费者泛在的信息环境, 易于与其他系统的集成。

从上述两个角度对目前基于功能的机构库系统解构后形成了基于服务的机构库框架。从系统角度的方面来看, 基于服务的框架是建立在基于功能的框架之上的, 因此两个框架是松耦合的, 不会互相影响。从实现的角度来看, 基于服务的机构库框架是调用了基于功能矿机的公共接口函数来实现的。基于服务的机构库把基于功能的机构库包装成了一种 Web 服务, 是第三方的系统可以利用服务公共接口函数来与机构库集成。基于服务的机构库中的服务功能的实现可以采用不同的方式来实现 Web 服务, 比如可以采用 SOAP 的方式, 也可以采用更轻量级的 REST 实现方式或者两者结合使用。

3.3 嵌入式应用部分

基于上述框架, IR 的核心功能都以 Web 服务的方式暴露了出来, 但这种方式仍然是一种可编程的应用程序接口, 系统之间通过调用接口函数来实现系统之间的互相集成, 在不同的系统之间交换数据, 实现资源的共享和互交换。对于普通的用户仍然没有办法直接地使用, 为了方便普通用户使用, 可以用以下的两种方式来实现用户级别的可嵌入。

(1) 用户桌面小工具

以基于服务的 IR 提供的功能为基础, 把需要的功能封装成一种桌面应用小工具。例如可以把资源的提交服务做成一种桌面的小工具, 用户要提交的资源是存储在本地, 因此这种方式非常适合资源的提交, 用户可以随时的提交资源而不用登陆到机构库系统中。

(2) JavaScript 的方式

在 Web2.0 的环境下, 强调个性化、互动性、平等性和细分化^[9]。基于 Web2.0

的理念网络形成了独特的应用,例如 Blog、RSS 等。用户的工作越来越依赖这些应用,逐渐的成为用户的信息环境的一部分。而这些系统大多数是通过 XML 来进行内容的共享和数据的交换的,因此把 IR 的服务以 Javascript HTML 的方式来与这些系统集成,实现机构库功能的嵌入。

4 结语

本文拟通过对机构库体系结构的分析,对机构库以功能为导向的架构进行解构,形成以服务为导向的机构库的框架。服务为导向的机构库系统提高了机构库的可扩展性,降低了系统模块之间的耦合性,使其更好的和其它系统进行整合和协同。重构之后的机构库,降低了模块之间的耦合度,可以独立构成应用。各个模块以 Web Services 的方式提供服务接口,使机构库的功能可以被很容易的嵌入到其他系统之中,增强了系统之间的沟通和互交换,更好的融入到用户的环境中。

参考文献

- [1] 罗曼. 影响信息环境变化的内外因素分析[J]. 图书与情报, 1994(1):3-11.
- [2] Crow.Raym.The Case for Institutional Repositories: A SPARC Position Paper. SPARC: Scholarly Publishing & Academic Resources Coalition, 2002 <http://www.arl.org/sparc/IR/>[检索日期 2008-5-22].
- [3] handle[EB/OL]. <http://www.handle.net/>. [检索日期 2008-9-13].
- [4] 张海涛, 郑晓惠, 张成昱. 数字图书馆互操作研究: Z39.50 和 OAI 协议的比较. 现代图书情报与技术, 2003(2):15-20.
- [5] Khalaf, Keller, Leymann. Business Processes for Web Services: Principles and Application[J]. IBM System Journal, 2006(4):32-35.
- [6] Gadget. [EB/OL]. <http://en.wikipedia.org/wiki/Gadget>. [检索日期 2008-10-2].
- [7] 李广建, 黄永文, 张丽. IR: 现状、体系结构和发展趋势[J]. 现代图书情报技术. 2006(2):236-240.
- [8] SRU: Search and Retrieve via URL (standards, Library of Congress) [EB/OL]. <http://www.loc.gov/standards/sru/index.html>. [检索时间 2008-07-08].
- [9] 田华, 鄢喜爱. WEB2.0: 图书馆信息服务模式转化的催化剂[J]. 图书馆, 2008(1):63-64.

