



图书馆 RSS 应用探索

Application Research of RSS in the Library

赵 阳 (清华大学图书馆 北京 100084)

[摘 要] 本文介绍了 RSS 的含义和优势,探索了图书馆自建 RSS 服务系统架构和技术路线。在此基础上,总结了 RSS 在图书馆的应用,包括:基于 OPAC 的 RSS 服务、电子期刊 RSS 服务、本馆自建资源 RSS 服务、免费学术资源 RSS 服务和聚合等,认为图书馆应积极开展与 RSS 相关的参考资源服务。

[关键词] RSS 公共 RSS 个性化 RSS RSS 检索

[中图分类号] G250.7 [文献标识码] B

[Abstract] This paper introduces the meanings and advantages of RSS and explores the architecture and technical methods of self-established RSS service systems. On these bases, the paper summarizes the applications of RSS in the libraries, including RSS based on OPAC, e-journal RSS, self-established resource RSS, free academic resources RSS and integration and so on. At last, the paper suggests that libraries should develop the reference services related with RSS.

[Key words] RSS; Public RSS; Personalized RSS; RSS search

1 RSS 含义与优势

RSS 有多种解释,它可以表示真正简单聚合(Really Simple Syndication)、丰富站点摘要(Rich Site Summary)或 RDF 站点摘要(RDF Site Summary)。本质上 RSS 是一种基于 XML 的信息内容描述、发布和聚合技术,通过支持标准的 RSS Feed(RSS 信息源)格式实现信息内容的聚合和订阅。与传统信息获取方式相比,RSS 具有如下优势:

信息过滤:用户可选择感兴趣主题的 RSS Feed 内容进行订阅,过滤掉不感兴趣或无用信息。

信息聚合:可对订阅的信息进行集中管理,实现相关信息聚合。

信息推送:能直接将更新的信息即时主动推送到用户桌面,使用户不必直接访问网站就能得到更新的内容。

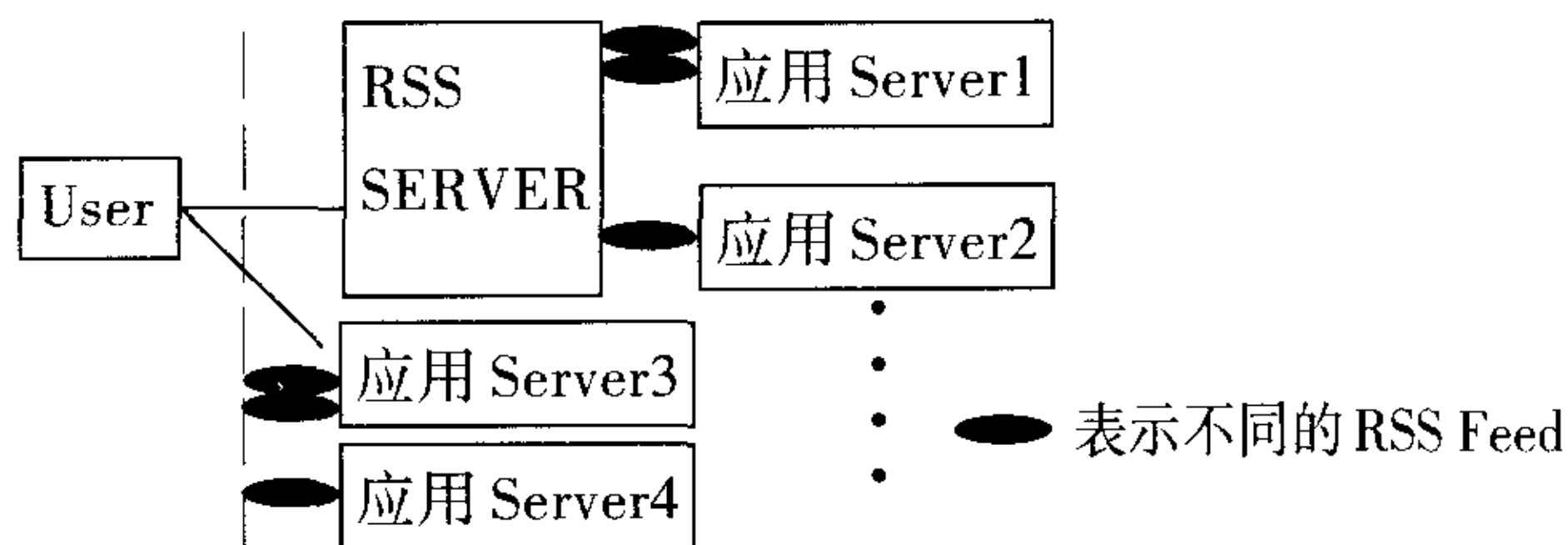
据统计,美国提供 RSS 内容的网站数目从 2001 年 9 月的 1 千余家增至 2004 年 9 月的 19 万 5 千余家,短短的 3 年中增长了近 150 倍。权威人士认为,未来 5 年内 RSS 将成为互联网中发展最快的技术之一。随着 RSS 技术的普及和广泛应用,国内外图书馆界开始积极尝试采用 RSS 来提高自己的服务水平。

2 图书馆自建 RSS 服务系统架构和技术路线

RSS 服务采用传统的客户端/服务器结构,系统架构可参看图 1,图书馆如需提供 RSS 服务,要构建 RSS 服务端。可根据应用服务系统的不同状况,选择不同的 RSS 服务方式。应用服务器可直接作为 RSS 服务方输出并发布标准格式的 RSS Feed,也可通过输出 RSS Server 支持的标准格式,由 RSS Server 间接提供 RSS 服务。同一个应用服务可根据

用户需要提供多个 RSS Feed。在用户端可以通过 RSS 阅读器订阅图书馆提供的 RSS 服务。具体地说:

图 1 图书馆 RSS 服务架构



2.1 在服务器端

首先服务器端需确定应用系统输出的 RSS Feed 格式和生成方式, RSS Feed 本质上是一个 XML 文件,由多个 XML 标签(tag)构成。一个 RSS Feed 包括两部分:网站的静态信息和网站中每一条目的动态信息。每一条目以<item>标签开始,包括 TITLE, Link 和 Description 等标签,并对应与之匹配的结束标签。根据应用系统的不同, RSS Feed 的生成方式大致有如下几类:

(1)手工生成 手写符合 RSS 标准的 XML 文件,通过 RSS Server 发布。该方式效率低,适用于内容不经常更新,包含 item 较少的 RSS Feed。

(2)RSS Editor 通过在 RSS editor 表单中输入相应的 item 信息,自动生成标准的 XML 文件。有许多基于 Web 的或可免费下载的 RSS 编辑器可供使用(如: RSS Creator, FeedSpring, FeedWrite)。

(3)从 HTML 文件生成 Feed 通过分析 HTML 标签批

量生成XML文件。有多种工具可供使用(如:基于Web的FeedFire、MyRSS,免费下载的RSS DreamFeeder、RSS Wizard)。

(4)基于数据库生成静态XML文件 当数据库中新增记录时,通过触发器技术自动生成静态的XML文件。该方式既保证了用户在数据更新的第一时间得到数据,也避免了因客户端频繁地与数据库交互而影响应用系统的性能。

(5)基于数据库动态提取数据 根据用户的请求动态生成XML文件,适用于个性化定制推送服务(Personalized RSS),比如基于关键词的个性化订阅,或者基于用户身份认证的RSS Feed。由于客户端定期向RSS服务端发请求,系统开销要大于前4种。

图书馆的数据大部分都通过数据库管理,后两种RSS Feed生成方式将被广泛使用。为便于集中管理和维护RSS Feed,静态XML文件(前4种方式)形式的RSS Feed可通过RSS Server集中发布、管理和提供服务。

服务器端通过支持OPML(Outline Processor Markup Language)可实现集中推送和管理多个RSS Feed。OPML是建立在XML基础上的“大纲处理标记语言”,用于描述RSS Feed列表信息,用户导入一个OPML文件,可以同时订阅多个RSS Feed。

2.2 在客户端

用户可通过RSS客户端工具,根据设定的时间间隔定期追踪服务器端RSS Feed的变化,订阅RSS服务有如下几种方式:

(1)RSS在线阅读 通过建立用户帐号的方式在线订阅和阅读RSS信息。与电子邮件的收发类似,对于并不总在固定地点上网,或者不需要另外安装软件来阅读RSS的用户适用。缺点是受服务稳定与速度快慢的影响。blogline、My Google等是比较有声望的RSS在线阅读网站,国内的RSS在线阅读服务刚刚起步,稳定性和功能都比较差。

(2)下载安装RSS阅读器 优点是用户拥有较高的控制权,速度快,方便搜索和管理自己的信息。缺点是本地安装软件占用一定的系统资源,移动性差,只能在软件运行并在线时才能更新RSS Feed,导致有可能错过一些信息。常见的有:sharp reader,周博通等^[1]。

(3)插件方式 通过安装电子邮件客户端或web browser插件,支持RSS阅读管理功能。该方式RSS阅读管理功能比较简单,节省系统资源,在浏览网页的同时可收集RSS Feed。不适用于对RSS的复杂管理。如:安装Pluck或sage插件在IE或Firefox上进行RSS阅读,安装Newsgtor插件与Microsoft Outlook集成订阅RSS服务。

(4)在移动设备上 可在手机或PDA等移动设备上,安装软件(如PalmRSS),来订阅RSS Feed。

选择RSS阅读方式关键取决于是否适合于自己的应用环境和偏好。

3 RSS在图书馆应用探索

图书馆拥有丰富的数字化资源,这为提供RSS服务提

供了资源保障。根据现有的资源状况,图书馆可提供3类的RSS服务:自建RSS Feed,已订购商业数据库提供的RSS Feed,免费的、有学术价值的RSS Feed。其中,自建RSS Feed是针对图书馆自建资源,或者本身不支持RSS服务但需迫切提供RSS服务的商用产品,需要一定程度的开发工作。对后两种RSS服务,参考馆员需要做一定程度的整合、筛选和推荐工作。RSS在图书馆的应用可归纳如下:

3.1 基于OPAC的RSS服务

基于OPAC提供RSS推送服务是未来的一个发展趋势,国外知名OPAC厂商(如INNOPAC, SIRSI, ALPHA等)均已声称在未来的系统升级版本中,支持RSS功能。图书馆如需在现阶段提供RSS服务,需要做一定程度的本地化工作。根据国内外的经验,基于OPAC的RSS服务可分为:

(1)新书通告

图书馆新书更新频率很快,定期将OPAC中新书主动推送给读者,能帮助读者追踪馆藏变化,提高服务水平。有两种实现方式:①通过Z39.50服务器动态读取OPAC中数据,按类输出标准的XML文件;②通过OPAC提供的数据输出接口,定期抓取符合条件的数据,编写程序自动进行数据格式转化,生成RSS Feed。根据不同的OPAC系统可选择不同的方式。清华图书馆新书通告RSS栏目,基于INNOPAC系统自编软件,实现了每周自动更新数据,按社会科学、经济学、法学、艺术、工程、综合等八大主题向读者提供新书RSS Feed,读者可按主题订阅自己感兴趣的新书通告,也可通过OPML连接一次订阅全部主题的新书书目。

(2)基于OPAC的个性化RSS服务^[2]

与新书通告RSS(属Public RSS)不同,个性化RSS基于用户身份提供个性化的RSS服务。订阅RSS Feed的URL包括用户的身份信息和个性化需求,可实现超期催还、超期罚款、预约书到馆通告、个人借阅信息列表等主动的RSS推送服务,方便读者了解自己的借阅状态,是传统方式下Email通告的一个重要补充和进步。技术上有两种实现方式:

●在OPAC系统中集成RSS功能:用户RSS订阅连接不直接显示在公共WEB页面,而显示在用户登录后的个人借阅情况查询页面,也就是说只有用户本人才能看到自己的RSS Feed URL。该方式需要了解OPAC系统底层结构,不适合做本地化开发。Tails、INNOPAC等OPAC系统将在升级版本中通过该方式提供RSS个性化服务。

●通过第三方接口实现从OPAC中提取个人借阅状况:为个性化RSS服务用户提供注册入口,必备的注册信息包括:卡号、PIN(来自OPAC用户库)和Email/Username/Password(存入自建的个性化RSS用户库),在后台通过页面分析技术或OPAC API接口,通过连接到OPAC用户库进行用户身份认证,得到用户的借阅信息,根据获得的信息构建个人的RSS Feed,将RSS Feed URL通过Email返回给用户,用户也可通过登记的Username和Password查询自己的RSS Feed URL。该方式原理简单,适合做本地化开发,也适合做扩展,满足其它应用系统的个性化RSS需求。ELF

(libraryelf.com)通过该方式实现个性化 RSS 服务。

个性化 RSS Feed URL 专属于某个人,除了用户保密自己的 RSS Feed URL,还需要从技术上实现安全性。目前的 RSS Feed 加密方法有:基于 SSL 证书加密(用 https 订阅)、存取令牌加密等、对 RSS Feed URL 中相关用户身份的部分进行加密,避免用户在发送 RSS 请求时明文的用户信息被截取等。

3.2 电子期刊 RSS 服务^[3]

目前很多数据库出版商提供 RSS 服务,主动向用户推送数据库资源变更信息,归纳起来有两类服务:

(1)基于检索词的订购 用户在检索页面中输入检索式,在返回检索列表的同时系统会自动给出一个 RSS Feed URL。用户订阅这个 URL 后,每次数据库更新记录的同时,用户将得到相关的最新信息。EI、Proquest、PubMed 等商业数据库提供该类服务。

(2)基于期刊目次(TOC)的订购 按照收录期刊的目次提供 RSS Feed URL,当新一期的目次入库时,自动通知用户最新的期刊目次信息。有代表性的商业数据库有:IEEE、IOP、Nature、Science、中国期刊网等。

图书馆可按照这个分类提供商业数据库 RSS 服务列表,以方便用户订阅。对于基于期刊目次的 RSS 服务,因揭示到单个期刊一级,如果本馆馆藏中提供 RSS 服务的期刊较少,可只提供简单的期刊列表和相应的 RSS Feed URL(如 Wisconsin-Madison 大学);如果馆藏中提供 RSS 服务的期刊较多,可与现有的电子期刊导航系统集成,方便用户按照期刊名查询和订阅 RSS 服务(如 Saskatchewan 大学)。由于许多商业数据库基于 IP 认证,MIT 为方便校园用户在校外订阅 RSS 服务,建立了图书馆 RSS 代理服务器,用户在订阅时只需要按要求改造 RSS Feed URL 即可。

3.3 本馆自建资源 RSS 服务

图书馆有许多的自建资源,按照资源的种类,可提供如下 3 类 RSS 服务:

(1)消息类 将图书馆的最新消息、公告和通知等通过 RSS 服务推送给用户。例如,MIT 提供的 MIT 图书馆新闻 RSS Feed,清华大学图书馆提供的图书馆公告(最新通知、培训信息等)、电子资源动态(数据库试用信息)、报告厅学术信息(学术报告信息)等 RSS Feed。

(2)资源类 将一些资源数据库中更新的最新数据推送给用户。例如:学位论文和电子教参系统 RSS 服务,武汉理工大学建立的学科导航系统 RSS 服务。

(3)Blog 类 国内外一些图书馆已经或正在尝试建立图书馆博客网站,发布及时消息,建立与读者良好的互动关系。Library Weblogs(www.libdex.com/weblogs.html)网站提供已有的图书馆博客网站列表,并不断更新。blog 软件可免费下载安装,并自带 RSS 功能,不需要另行开发。

对于何种自建资源需提供 RSS 服务,最好成立一个讨论小组结合读者需求来确定。一般来说,资源内容更新频次较大的应用适合创建 RSS 服务。在技术上,图书馆自建系统是有共性的,一般都是基于 ASP、JSP、PHP 等动态发

布技术,以 Access、Oracle、MySQL、SQL-server 等作为通用的后台数据库。因此,在 RSS 服务技术实现上具有共通性,按平台类型整理出一个开发手册,可提高开发效率节省时间。对于 Public RSS 建议采用 2.1 小节中的方法(4)来开发,对减少应用系统负担;对 Personalized RSS 或基于检索词的 RSS 订阅,可采用 2.1 小节中方法(5)来开发。

3.4 免费学术资源 RSS 服务和聚合

目前,网络上许多免费学术站点也支持 RSS Feed 输出。对这些资源进行整理、筛选和归类,推荐给读者使用是非常有价值的。例如:Dlib、Ariadne 等图书馆界知名的网络期刊支持 RSS 服务,CiteULike 提供学术论文、专著的 RSS Feed 输出,有许多专业性、学术性较强的开放存取期刊也支持 RSS。这项工作类似于网络资源整理和推荐,只不过对象是 RSS Feed,学科馆员凭借自己对专业信息资源的了解和学科背景,是开展这项工作的最佳人选。

一些图书馆也在尝试建立本馆 RSS 信息聚合系统,人工选择网络上较关注主题的、有价值的 RSS Feed,通过软件自动实现 RSS Feed 实时或定时地收集到本地的 RSS 信息数据库中,按照一定的组织方法对收集的 RSS 信息进行统一的整合、组织、揭示和管理,并对其提供统一的检索、浏览和 RSS 输出等增值服务功能。有代表性的如:加拿大肿瘤学院图书馆建立的 RSS 信息聚合系统 LION(Library Online),中国科学院文献情报中心建立的“基于 RSS 的科技信息聚合系统”等。图书馆可使用一些免费的开放源码软件(如 ROSA、Urchin 等)建立这样的系统。

4 结 语

作为 library2.0 重要组成的 RSS 技术,将为图书馆提供新的服务思路 and 手段,为用户获取信息提供了更多的选择。随着 RSS 的广泛应用,将来会有更多的资源提供者支持 RSS 技术,互联网上将存在浩如烟海的 RSS 资源,同时,搜索业巨头 Google、Yahoo 等已捷足先登,新近推出了 My Google、My Yahoo 等功能强大的 RSS 个性化服务,Microsoft 宣称将在新版 IE 中支持 RSS 功能,这些可能会吸引更多的图书馆用户使用其新服务而远离图书馆,图书馆面临更多新技术的挑战。在新的发展机会和挑战并存的情况下,相信图书馆员凭借其组织、加工和管理信息资源的特长,凭借新技术和创新思想做依托,将会为用户提供更加专业的、高层次的信息服务。

参考文献:

- 1 Fudith Wusteman. RSS: the latest feed. Library Hi Tech, 2004(4): 404-413
- 2 Richard Wallis. Talis RSS Whitepaper. http://www.talis.com/research/research/rss/rss_whitepaper.pdf, 2006.03.05
- 3 Zeki Celibas. What is RSS and how can it serve libraries. http://eprints.rclis.org/archive/00002531/01/RSS_and_libraries_EN3.pdf, 2006.3.8

[作者简介]

赵 阳 清华大学图书馆系统部工作,中级职称,发表论文 10 余篇。
[收稿日期:2006-04-24]