

开放链接技术研究

李春晓 梁康 左春 (中国科学技术大学、中科院软件研究所)

摘要: 本文依托数字图书馆的建设背景, 阐明了数字资源建设中的数字资源基于内容的相关性, 重点研究了当前倍受业界关注的解决基于内容链接的技术标准——开放链接 (OpenURL) 的基本概念、工作流程及其应用模式。

关键词: 数字图书馆 数字资源 开放链接 (OpenURL) 来源资源 目标资源

1 引言

随着数字信息资源、信息服务系统和用户信息环境的不断发展, 数字图书馆机制也从基于数字信息资源的系统形态 (第一代) 逐步过渡到基于集成信息服务 (第二代) 的系统形态, 并开始向基于用户信息活动环境 (第三代) 的系统形态过渡。

因此, 数字图书馆不再仅仅是一个基于资源的服务系统, 而是一种以用户为中心来聚合资源、服务、信息利用活动的动态机制, 其目标和功能都着力于支持用户利用信息、提炼知识、解决问题, 更多地直接面向用户对信息进行感知、捕获、分析、重组、传递、应用的全过程及其多样化需求, 通过资源与工具的链接、集成和嵌入来动态构建恰好符合用户需要的服务机制。

从第三代数字图书馆角度, 第一代、第二代数字图书馆系统都应是能被灵活和动态地嵌入到用户信息处理环境中的信息资源与服务机制, 这就要求这些资源和服务机制在更广泛基础上实现标准化、开放性和互操作, 支持基于内容和知识组织体系的开放性的信息析取、转换、注解、链接、重组和传递, 支持信息内容、资源系统和信息服务过程的开放性定义、描述、解析、链接、构建、嵌入和集成, 支持对信息内容、功能模块

和使用机制的可伸缩和分布式管理控制, 从而保障数字图书馆系统及其任意内容或功能部分能被灵活、有机、可靠和可控制地嵌入用户信息系统和用户信息活动过程中。

在数字图书馆的基础——数字资源的建设和利用中, 我们注意到数字资源之间存在的基于知识和内容的复杂关联关系。

大多数较有参考价值和权威性的数字资源一般以数据库的形式存在, 比如 ISI 的 Web of Science, EI 的工程索引, EBSCO 的 CC, Pubmed, Science 等。

一般地, 这些数字资源库可以按照不同的标准进行划分:

(1) 按照数字资源存在的区域划分:

内部网数字资源库

因特网数字资源库

(2) 按照数字资源本身的内容划分:

文摘数据库

全文数据库

引文数据库

联合目录数据库

OPAC 等

(3) 按照数字资源的建设者划分:

机构自建的特色资源库

商用数字资源库等

这些数字资源库由于建设者不同, 导致平台异构, 使用方法、覆盖范围不尽相同。但是, 文献之间的天然联系却不容忽视。以期刊为例, 可能某种期刊同时被不同的数字资源库收录, 这些数字资源库或迅速展示目录 (current contents), 或记录文摘信息, 或制作权威学术标引, 或提供全文传递服务……等等, 不同形式描述同一文献资源。不仅如此, 文献之间还可能存在着引用和被引用的关系, 按照不同的相关标准 (主题相关或是共引相关), 还可能存在着不同程度的相关关系。总之, 数字文献资源之间存在着千丝万缕的基于内容和知识的相互关联。

这种内容关联, 反映在基于 web 的应用上,

就是各种形式的链接关系——参考链接、引文链接、相关链接、服务链接、逻辑链接等等。

那么, DL 建设者们如何能够保证, 用户在异构且相关的数字资源中自由有序地穿行, 到最适合的位置获取他有权获取的数字资源 (appropriate copy), 而不至于迷失方向, 或者为低效率的工作所困惑呢?

这就是开放链接技术期望解决的问题。

通过开放链接技术, 将数字资源的内容进行无缝链接, 并对所链接对象进行本地化控制, 支持可定制的一站式检索, 支持用户对多种信息内容的集成利用。

正如美国国家信息标准组织标准委员会 (NISO) 所说: "At Openly Informatics, we recognize that integrating new linking services is a complex and difficult task for librarians. Users need consistent interfaces and should be shielded from the complexities of authentication and control. Disparate systems need patching together. And in the end, things have to work. That's why we're here to help." (关于开放式情报, 我们认识到整合集成链接服务对于图书馆员来说是一项复杂和困难的工作, 而用户需要一致的界面并且应该屏蔽掉复杂的授权和控制; 同时, 各种不同的系统需要联合。无论如何, 事情总归要做, 这就是我们在这儿提供帮助的原因。)

2 开放链接 (OpenURL) 技术

我们可以从了解开放链接的研发现状、基本概念、工作流程和 OpenURL 语法入手, 得出与传统链接相较的优势, 从而证明引入开放链接的必要性, 进而讨论可能的应用模式和扩展范围。

2.1 开放链接的研发现状

开放链接 (OpenURL) 是美国国家信息标准组织标准委员会 (NISO) 2001 年 5 月提出草案建议的一个正在讨论确定的标准。它得到了世界范围的广泛响应和支持 (如 Web of Science, ProQuest 等), 国外的一些科研机构和公司 (如: ISI-Web of Science, OCLC-First

The Research of OpenURL Technology

Search)对OpenURL技术已进行了深入的研究和开发,使开放链接技术正在成为一项趋于成熟的技术。数家公司(如Exlibris,Endeavor)也已经推出了相应的产品。仅Exlibris的产品(SFX)第一年就拥有了世界范围的70个用户。虽然标准本身还有需要界定、确认的诸多问题,但是它的受欢迎程度是毋庸置疑的。国内正在从事开放链接技术的研究还不多,而业界仍然没有成熟的产品推出。因此,国内在相关产品方面还是一个空白。

2.2 开放链接基本概念

OpenURL即开放链接,是一种基于Web,使用指定语法传送元数据的标准,它用来为元数据和(或)一个信息对象的标识符创建web传输包,这样的包是上下文相关链接或开放链接技术的核心。通过标准化的语法,OpenURL可以激活许多其他新的、特定用户的服务。由于具有可定制、可移植、可扩展的开放特性,已逐渐成为数字图书馆应用领域的热门技术。OpenURL标准是构建开放式数字图书馆的关键组成部分之一。

相关的重要概念有来源资源、目标资源、适当拷贝、扩展服务等。

来源资源(Source):Source是在链接检索中

用来显示给用户的信息,这个信息可能是一个引用或者是来自于商业数据库的摘要、OPAC中的一个书目提要信息、OPAC中一个特别卷的项目信息、标引的全文或其他任一项目,它们可以生成结构化元数据并且支持OpenURL标准。

目标资源(Target):Target是一个资源的定位,link-to语句可以从这个定位里检索到信息。目标资源可能包含一个全文数据库、OPAC,引文或文摘数据库和其他可扩展的服务。

适当拷贝(Appropriate copy):它指一个可用资源(如文献等)的“最好的”由机构定义的拷贝。一般来说,适当拷贝就是指通过最少的开销就可以获得的拷贝。

扩展服务(Extended Service):提供一些附加的、相关联的服务,比如全文发送服务,通过本地的OPAC查询,或者在有兴趣的站点进行搜索等等。

开放链接的特点是具有动态特性,其关键在于,将来源资源中的链接,用标准开放的方式描述——符合OpenURL语法规则,因而引入了开放链接解析服务器,通过解析标准的OpenURL请求,以扩展服务的形式,间接定位到目标资源的适当拷贝。

2.3 开放链接工作流程(如图1)

开放链接的动态特性,是以由其特有的工作流程决定的。

2.4 OpenURL 语法

开放链接工作流程中的OpenURL请求,必须符合OpenURL语法,目前定义的标准OpenURL语法,如下所示。

OpenURL主体包括两部分:

(1)基础URL(BASE-URL):OpenURL解析服务器的URL地址

(2)查询(QUERY):对象源数据及其提供者描述,分作源标识区和对象元数据区,其中:

①源标识区(ORIGIN-DESCRIPTION)标识数据库及其供应商;

②对象元数据区(OBJECT-DESCRIPTION)描述对象的元数据或唯一标识符(包括统一唯一表示符和本地唯一表示符),其中元数据方式最为直接。

下面是一个元数据表示的OpenURL示例,表明来源资源的对象是一篇文章,其母体文献的issn号为1234-5678,刊登在第12卷第3期的第1页到第8页,其作者是Paul.Smith,1998年发表。
`http://sfx.aaa.edu/menugenre=article&issn=12345678&volume=12&issue=3&spage=1&epage=8&date=1998&au1ast=Smith&au1first`

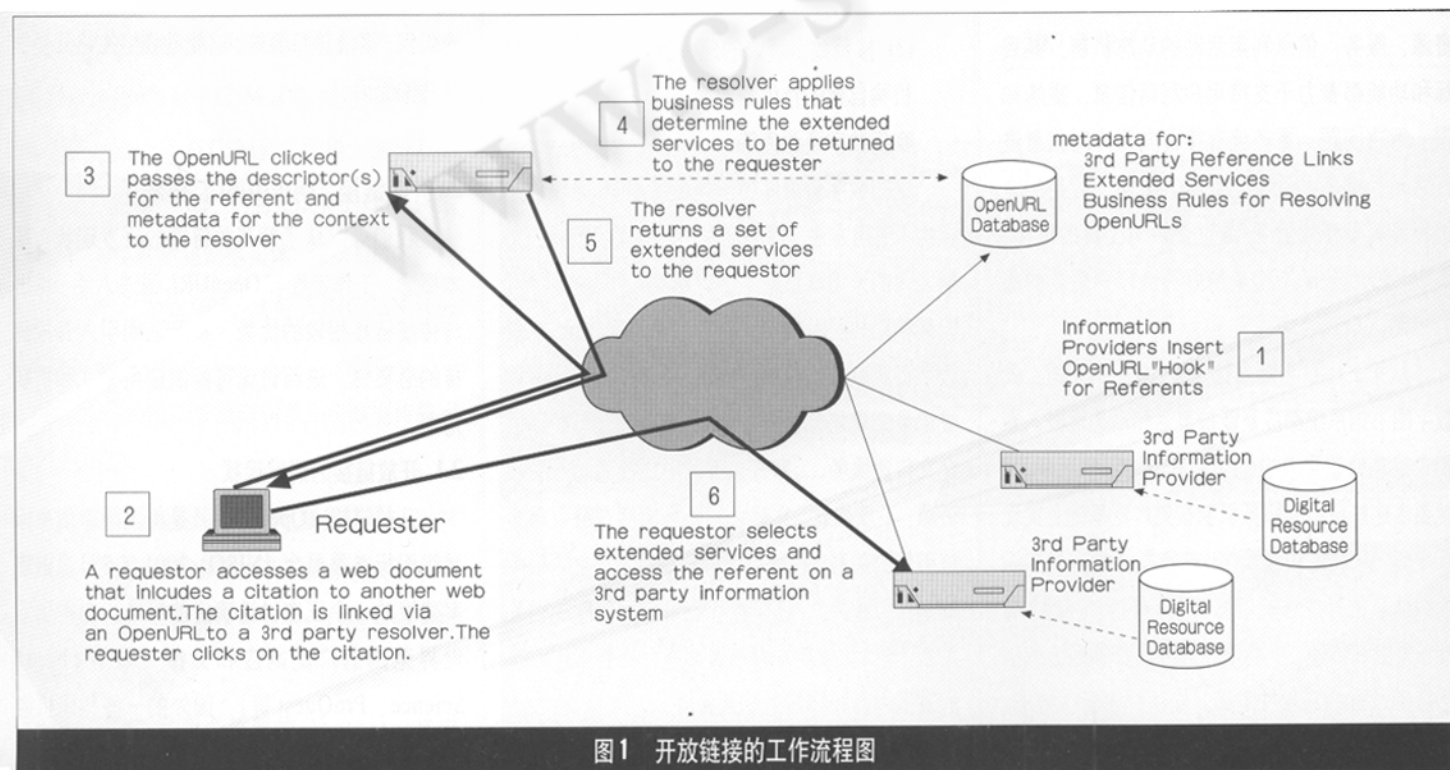


图1 开放链接的工作流程图

步骤	描述
1	数字资源提供者为了来自源中不同的实体向解析服务器插入一个 OpenURL "hook", 解析服务器负责接收和处理 OpenURL 请求。
2	请求者点击 OpenURL 链接, 通过发送 OpenURL 请求访问解析服务器。
3	OpenURL 请求传输实体元数据和相关信息给解析服务器。
4	解析服务器根据一些商业规则 (这些商业规则可由本地化管理) 返回对这个 OpenURL 的回答。
5	解析服务器返回一系列服务类型 (扩展服务菜单) 给请求者。
6	请求者通过一个或多个服务类型访问数字化资源。

=Paul

2.5 引入开放链接的必要性

由上述例子, 可以看出开放链接与常规链接的不同。

所谓常规链接是针对开放链接而言的, 基本特征是链接直接写成固定的 URL, 用户点击的结果是直接指向目标。

如 http://ojps.aip.org/journal_cgi/dbt?KEY=JCP&volume=111&Issue=20&sp=9398, 因此, 常规链接不可避免地存在着下述限制: 常规链接由来源数据控制, 是静态的链接, 是上下文不相关联的; 常规链接由来源数据提供方定义, 无法实现机构端本地化控制; 常规链接更新和维护比较困难, 链接对象的细微变化都会引起链接的失效。

对应地, 开放链接允许动态链接, 提供上下文相关联 (上下文敏感, Context Sensitive) 的链

接; 可以实现对链接对象进行本地化控制 (提供合适的链接); 开放链接为所有的潜在数据源提供单点的服务管理, 开放链接能提供广范围的服务 (提供不同的链接), 实现资源和服务的扩展; 开放链接允许机构的管理员定制服务; 基于 OpenURL 标准, 保证链接的规范性和标准化。

开放链接与传统链接相较的优势, 说明了引入开放链接的必要性。

2.6 开放链接的应用

目前, 开放链接的典型应用是情报机构, 因为:

(1) 随着文献资源数字化建设的投入加大, 为用户提供的电子文献资源的种类和数量也随之快速增长。

(2) 数字化资源虽然源自不同的提供者 (购买、自建、其他或第三方), 平台各异, 但在内容上却存在着复杂的交叉关系网。

(3) 资料在不断成长的情况下, 使得资源的

定址应用出现困扰, 常规链接的更新和维日益复杂和繁乱。

(4) 用户希望统一、开放、友好的解决方案。

(5) OpenURL 标准提供改善情况的契机。

(6) 情报机构 (图书馆) 直接面对资源和用户, 是二者联系的桥梁, 也是本平台搭建的理想基地。

(7) 情报机构随时可以得到服务情况的改善报告、用户问题报告, 以便持续改善服务。

所以, 开放链接平台 (核心是 OpenURL 解析服务器) 本身要求具有良好的可扩展性, 应以情报机构为中心搭建, 从已经支持 OpenURL 的资源出发, 逐步扩展整合到所有资源, 同时建立与国际通用的数字对象唯一标识符解析系统的互操作机制。

3 结束语

开放链接技术平台的研究与建设, 在数字图书馆的资源建设和利用中扮演着重要的角色, 它为解决数字资源基于内容的相关性问题, 提供了值得广泛推广和应用的解决方案。 ■

参考文献

- 1 张晓林, 数字图书馆机制的范式演变及其挑战, 情报学报, 2001。
- 2 郑巧英, 杨宗英, 中国数字图书馆试验基地实例, 现代图书情报技术, 年刊/2001: 21-23。
- 3 徐文伯, 中国数字图书馆, 数字资源的开发与研究, 人民出版社, 2002/5。
- 4 张晓林, 开放数字环境下的参考链接, 现代图书情报技术, 2002/1: 9-13。

```

OpenURL::=BASE-URL'?QUERY
QUERY::=DESCRIPTION('&&DESCRIPTION)
DESCRIPTION::=(ORIGIN-DESCRIPTION'&')? OBJECT-DESCRIPTION|
OBJECT-DESCRIPTION('&'ORIGIN-DESCRIPTION)?
ORIGIN-DESCRIPTION::=sid='VendorID'"DatabaseID
OBJECT-DESCRIPTION::=ZONE('&'ZONE)*
ZONE::=(GLOBAL-IDENTIFIER-ZONE|OBJECT-METADATA-ZONE|LOCAL-IDENTIFIER-ZONE)
GLOBAL-IDENTIFIER-ZONE::='id' '=' GLOBAL-NAMESPACE ':' GLOBAL-IDENTIFIER
('&' id' '=' GLOBAL-NAMESPACE ':' GLOBAL-IDENTIFIER)*
GLOBAL-NAMESPACE::=('oi' | 'pmid' | 'bibcode' | 'oai')
GLOBAL-IDENTIFIER::=VCHAR+
OBJECT-METADATA-ZONE::=META-TAG '=' META-VALUE(& META-TAG '=' META-VALUE)*
META-TAG::=('genre' | 'auiast' | 'auiirst' | 'auiunit1' | 'auiunitm' | 'coden' | 'issn' | 'eissn' | 'isbn' | 'title'
| 'stitle' | 'atitle' | 'volume' | 'part' | 'issue' | 'spage' | 'epage' | 'pages' | 'artnum' | 'sici' | 'bici' | 'ssn' | 'quarter' | 'date')
META-VALUE::=VCHAR+
LOCAL-IDENTIFIER-ZONE::='pid' '=' VCHAR+
    
```