●王艳翠 编译（聊城大学 图书馆，山东 聊城 252059）

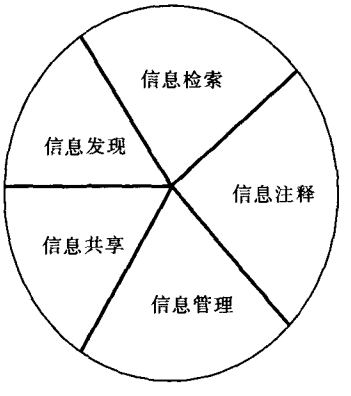
斯坦福大学数字图书馆技术

[关键词] 斯坦福大学；数字图书馆技术

[摘 要] 介绍了斯坦福大学数字图书馆项目的主要内容，包括信息检索的共用协议、价值过滤、网页聚类的技术，信息管理中的档案式存贮、InterBib、PDA 图片浏览技术，信息共享中的 DietORB、数字钱包、移动接入技术等一系列数字图书馆新技术。

[中图分类号] G250.76；G259.712
[文献标志码] B
[文章编号] 1005－8214(2007)05－0103－02

斯坦福大学于 2004 年结束了 DLI2（Digitial Library Technologies Project）阶段的数字图书馆技术项目。这个项目是于 1999 年在政府、大学和合作者的支持下开始的。项目的目的是设计和实施数字图书馆中由于信息资源的合作开发、传播、共享和管理而需要的基础设施和服务。下图是斯坦福大学数字图书馆项目的计划。



建立一个好的数字图书馆包含的不仅是研究和发展那些获取信息的工具。或许在数字图书馆中有许多信息源，而每个信息源中又包含了大量的内容，那么哪一个是我们要获取的信息呢？一旦用户获取到了多个信息源，我们能否为用户提供工具来帮助用户解释和处理查询结果，并使用户同样地为数字图书馆的服务付费吗？一旦用户获取到了他们所需要的信息，他们怎样才能高效地与他人共享呢？斯坦福大学数字图书馆研究涉及到了上述问题的方方面面，并且在数字图书馆的次级项目中解决了部分问题。

1 检索信息

检索某一特定信息源并把结果反馈给用户。

• 斯坦福数字图书馆 PalmPilot（手持电脑）基础设施：提供容错、事件日志、内存管理，并通过通信基础设施为发展迅速的个人数字助理（PDA）在数字图书馆中的应用提供帮助。

• 超强浏览。超强浏览允许巴掌大小的计算机浏览互联网。小如 PalmPilot 的屏幕，带宽如同便宜的无线电链接——要求完全重新考虑用户界面对信息库（例如互联网）的浏览。这个项目开发了浏览的新方法，它使用了多个我们研发的能使浏览方便快捷的支持设施。其中一个是导航，使用户的浏览速度可以提高 45%，另一个提供动态的网站查询和自动根据主题词词条完成查询。另外一项技术是帮助用户在他们的小设备上分析单个网页，这个技术提供语法和结构性的网页摘要以及逐步揭示网页内容的渐进机制。

• SDLIP（Simple Digital Library Interoperability）：简单的数字图书馆共用协议。

简单的数字图书馆共用协议是一个集中多个不同的信息资源库的协议。它是由斯坦福大学、加州伯克利分校、加州圣巴普拉、圣地亚哥超级计算机中心和加利福尼亚数字图书馆项目联合开发的。用户通过 SDLIP 协议来请求查询资源库，查询的结果同步返回，或是文件可利用的话，由服务器直接传送给用户。这样就可以构建基于 HTTP 或 CORBA 的传输。事实上，任何一个查询服务都可以同时通过这两种传输方式来执行。

• 问题翻译器。帮助用户查找支持不同查询语言的各类信息服务。这种方法允许用户自始至终使用统一的语言编制布尔逻辑查询，并把它们按照句法和功能转化成本来的形式。

• 价值过滤。价值过滤解决了搜索引擎超载以及搜索不到多媒体网页上的要素的问题。它是根据“文件价值”而不仅仅是根据“查询 / 文本”的相似性来搜索和过滤文件的。它通过有价值的信息来提高用户与信息之间的互动。

• 网页库（Webbase）。网页库项目探索大量的网页怎样能被高效地收集、存贮、检索和开发利用。斯坦福大学建立了巧妙的智能搜索器（Smart Crawlers），并建立了可存贮网上获取网页的存贮系统。网页库是研究人员建立的独特的入网索引工具。研究人员能够以非常高的速率通过系统为特征分析程序提供相应的网页，针对这些计算机化了

的网页特征，网页库就会建立专门的索引，这些索引随后将被用来进行查询。

• 网页聚合（web clustering）。在网页相关的查询中，一个艰巨的挑战是：网页的聚合与分类。网页聚合是指以一种与雅虎或是开放目录相似的方式，把网页编组分人有关各类。然而，这两个目录没有使用任何自动化技术，而是完全由人来编辑、维护。对于整个网来说，手工技术是不可升级的，尽管在网络（http://www.inktomi.com/webmap/）上有一万亿个网页（雅虎和开放目录在他们各自的体系内各有不到200亿个urls地址）。由于网页的庞大规模和超链接的性质，传统的IR方法在网络的上下文链接中是不恰当的。斯坦福大学最近开发了允许在高维度空间中进行相似性查询的技术，特别是如同精确度要求被提升了一样，即使有了更新的技术，信息资源的需求也会很大。在网页聚类和其他揭示网页内容的操作方面，高度计算机化的资源将是非常有价值的财产。这些资源将允许我们探索和评估更多可利用的聚类选择，如同我们开发最有效的技术。

2　信息管理（Managing information）

信息管理项目致力解决现实中信息管理的一些问题，包括：长期信息组织、分布式环境中的信息付费、版权侵权管理等。

• 档案式存贮。数字图书馆存贮库由相互独立但又相互合作的站点组成。每个站点管理着一些数字化资源，并对其他站点（已被定义了的）提供服务。

• 与参考书目相关的服务（InterBib）。InterBib是用来维护书目信息的工具。作为统一标准的、可查询的书目仓库，它能以许多不同的格式进行读写。

• PDA图片浏览器。随着个人数字助理（PDAs）的计算能力和存贮容量的增长，图片浏览器作为这些设备的可行和重要的应用而出现了，斯坦福大学研发了两种浏览器来支持PDAs中大量图片的收集。一种浏览器使用了一种传统的、文件夹式的布局：使用用户手动创建的组织结构或是系统自动生成的结构。另一种浏览器使用了基于垂直、可放大的时间轴的新颖接口。这个时间轴浏览器不要求用户组织它们的图片，反而是单纯依靠系统自动生成的结构。系统创建了一个用户图片的等级结构，它是基于用户图片的申请时间来聚类辨认可能相关的图片子集。在用户实验中，根据每位用户的图片收集来比较用户通过浏览器查找和浏览的结果；图片收集规模在500—3 000之间。结果表明，时间轴浏览器与传统浏览器在执行查找和浏览工作方面至少是同样有效的，而传统浏览器要求用户手工组织它们的图片。

3　信息共享（Sharing information）

信息共享工具包括：文件解释、用户界面和为视觉障碍者提供音像录入。

• Diet ORB。Diet ORB是高度缩小了的CORBA的掌上设备。斯坦福大学为个人数字助理的掌上电脑开发了一种CORBA ORB。ORB目前只允许利用个人数字助理来进行全方位的服务。

• 数字钱包。数字钱包的研究工作集中于实现电子支付场所的互通。简单的数字钱包结构包括：支付、交换、充值和其它操作。数字钱包是一个允许用户使用金融设施（如信用卡、数字货币）来进行电子支付的一个软件的组成部分，它省略了执行用以支付的付款协议的细节。

• 移动安全（Mobile Security）。图书馆的移动接入增加了经济基础设施的复杂性。为了达到移动接入的目的，用户的数字式证件可以由几台机器共享（家用电脑、笔记本和个人数字助理），相应地，所有机器在图书馆的相互作用中必须保持其一致性。例如，已在家用电脑上花费了的数字货币不应该再在笔记本上又花费一次。

[编译者简介] 王艳翠，女，山东大学信息学专业毕业，管理学在读硕士，现为聊城大学图书馆馆员，已发文数篇。

[收稿日期] 2006－10－24　　[责任编辑] 陈永平

动态·资料

宁夏初步实现“三网合一”共建共享

近日，记者从宁夏回族自治区党委、政府召开的全区新农村信息化建设现场会获悉，宁夏初步实现了网络、电话、电视的“三网合一”，基本实现了全区文化信息资源共享工程的应用功能。

目前，宁夏各市、县（区）全部建立了新农村信息化管理和服务机构，在所有乡镇配备了专职或兼职农村信息员，全区信息员已达3 600人。全区的187个乡镇实现了通光缆、通宽带，农村电视普及率达到95%以上，广播综合人口覆盖率达到98%以上，农村中学都拥有计算机教室，农村小学全部建成现代远程教育系统。全区初步实现了农村党员干部现代远程教育、文化信息资源共享、数字图书馆、农村中小学远程教育、网络电视、互联网经营等应用功能，初步达到了多网融合、共建共享的目的。

据了解，宁夏将充分利用“三网合一”，至2008年10月将使文化信息资源共享工程在全区2 382个行政村实现“村村通”。

——摘自2007年9月11日《中国文化报》