

焦玉英 索传军 张学红

论图书馆专业信息服务频道及其实现方法^{*}

摘要 图书馆专业信息服务频道可以理解为:为主动提供某个主题或专题信息相对集中的 Web 网页或 Web 站点。它包括信息的采集与加工,频道的定义和内容的组织管理,用户需求的获取与推送。其技术实现方法,一是利用频道生成工具建立,二是编写频道定义文件。参考文献 6。

关键词 网站 频道技术 图书馆 信息服务

分类号 G250.76

ABSTRACT Library subject information service channel is a Web page or site providing information on a special subject. It includes information acquisition and processing, channel definition and content management, etc. Its techniques include the creation with channel generation tools and the programming of channel definition files. 6 refs.

KEY WORDS Website. Channel technique. Library. Information service.

CLASS NUMBER G250.76

1 专业信息服务频道概述

所谓频道是指无线电和广播设备工作时,占用一定频率范围以传送电信号的通道,也称波道。频道广播是推(Push)技术实现方式之一。因特网上的频道是一个集中了某类信息,定期更新、定期通知的 Web 页。图书馆专业信息服务频道可以理解为:为主动提供某个主题或专题信息相对集中的 Web 网页或 Web 站点。如经济学信息服务频道,就是一个集中了经济学信息的 Web 页。

传统图书馆的文献信息服务通常是等待读者上门查询有关文献信息。如果图书馆依据某种方法,对馆藏(包括虚拟馆藏)信息进行挖掘、加工和重组,按照用户需要的方式设置成各类专业信息频道,供用户订阅,并将用户订阅的有关信息主动推送给他们,使用户不必到馆或重复查询有关数据库,就能获得自己所需的信息,必将提高广大用户信息获取和利用能力,提高馆藏文献信息的利用率,促进图书馆网络信息服务的层次。

2 专业信息服务频道工作原理与信息采集

2.1 信息服务频道的工作原理

一个频道站点要把用户所需的信息不断地推送给用户,必须解决“推送什么、如何推送、推送给谁”这 3 个问题。一个频道站点通常包括三部分:信息的采集与加工,频道的定义与内容的组织管理,用户需求的获取与推送。信息搜寻智能代理按照所设频道,自动搜寻有关信息,及时传回频道站点进行加工整理后存储到数据库中。信息推送代理定期监视搜寻代理数据库中数据的变化情况,定期更新有关频道内容。用户智能代理一方面随时监听用户的信息需求,管理用户需求特征数据库,另一方面根据用户的要求将其订阅的有关频道信息及时地传送给用户。

2.2 信息的采集

在专业信息服务频道建设中,信息的采集是频道服务面临的首要问题。对图书馆来说,用户所需的信息可能存在于馆藏中,也可能存在于馆藏外,如因特网上。如何有效地将用户所需的信息搜集起来,是保证服务的关键。对馆藏的各类数据库,可以利用数据库检索、数据挖掘技术、信息过滤技术等将有关信息或元信息(如书目信息等)集中起来。而基

参考文献

- 王大可. 数字图书馆. 深圳:海天出版社,2002
- 刘家真. 馆藏文献数字化的原则与方法(上). 中国图书馆学报,2001(5)
- 张萌. 图书馆文献数据库建设浅析. 津图学刊,1999(4)

- 吴利薇. 东北文献图录数据库与文献资源数字化建设. 图书馆学刊,1999(4)

何江涛 深圳图书馆馆员. 通讯地址:深圳市. 邮编 518027. (来稿时间:2002-04-01)

^{*}本文为国家自然科学基金项目(70073022)的研究成果之一。

于因特网的信息采集问题、信息的自动搜集与智能化加工等是难度很大的工作。尽管现在一些门户网站正在开展这项工作,但由于搜索的广度和加工的深度不够,查询效果较差。要真正实现因特网上信息的自动采集,必须将信息搜索技术、基于 Web 的数据挖掘技术、人工智能技术、知识发现技术等结合起来。

3 专业信息服务频道的设置

3.1 网站的频道设置方法

一些网站将某类信息集中存放在一个 Web 页面上,称为某某频道。如商都信息港(<http://www.zz.ha.cn>),在主页上设置了新闻频道(news.shangdu.com)、经济频道(economy.shangdu.com)等。在新闻频道下面,又设立了国内新闻、国际新闻和社会新闻等子频道。在子频道中存放着当天的新闻信息的标题。一些网站为了提高网站的点击率,通常根据网民的爱好设立频道和子频道。频道设置没有科学规律,存在很大的自主性。

3.2 图书馆专业信息服务频道的设置原则

图书馆的网络信息服务,更加强调信息服务的专业化、系统化和科学化,在设置专业信息服务频道时,对信息服务频道的规划要注重科学性和系统性,要保持频道和子频道之间的内在的科学联系;频道的命名或定义要既专业化,又要通俗易懂便于用户理解;要便于信息的组织和管理;要依据本馆馆藏和虚拟馆藏特点,有利于馆藏文献信息资源的开发利用;要面向本馆用户,要保证对用户个性化和多样化信息需求的满足。

3.3 图书馆专业信息服务频道的设置方法

图书馆在频道设置时,既要考虑图书馆文献信息的加工管理特点,又可以借鉴电台、电视台和频道网站采用基于主题的信息组织方式。频道可以依据《中国图书馆分类法》的22个大类,根据需要来选择,子频道可以依据用户的需求特点和习惯参照大类下的小类来选择。比如某大学有经贸学院,有政治经济学、工业经济、国际贸易、国际金融等专业,可以这样设立图书馆的经济信息频道:经济频道,下设政治经济学、工业经济、国际贸易、国际金融几个子频道。还可以根据信息服务项目的类型,在图书馆 Web 站点上设立新书通报频道、网络信息频道、专题信息频道等。

4 频道内容的组织

图书馆能否建设好频道站点,能否开展理想的

信息服务,满足用户个性化的信息需要,关键是频道内容的组织。

广播电视和网站都是大众媒体,在频道设置、内容组织和传播上,比较注重时间效应,在频道栏目(子频道)的组织上多以时间为主轴,是一种线形的组织方式。每个子频道的内容是一个相对集中的专题信息集(排序缺乏规律性),频道内容的组织方法较为简单。

图书馆不仅数据格式较为复杂,而且用户的信息需求也日趋多样。因而在频道内容组织上可采用不同的频道、不同的服务方式和不同的组织策略。对于大量的数字化内容,一般可以采用元数据库和数字对象数据库的组织存放方式。通常频道只显示或推送的元信息(关于信息的信息),当用户需要某条信息的全文时,再从对象数据库中调阅全文信息。

对信息的加工,比较适合采用 XML(eXtensible Markup Language,可扩展置标语言)来描述。使用 XML 中的 DTD(Document Type Definition,文档类型定义)或 XML Schema(XML 模式)来定义 XML 文件的逻辑结构,定义 XML 文件中元素、元素的属性以及元素和元素属性之间的关系。在内容显示时,可用 XSL(eXtensible Stylesheet Language,可扩展样式语言)来显示用户的个性化需求。需要说明的是,XSL 是表达样式表的语言,而所谓的样式,包括两方面的含义:首先是内容的重新组织,然后是内容的呈现格式。这种信息内容的组织方式,突出的优点是可以实现数据内容存放和数据显示样式相分离,可以很容易地满足不同用户的不同订制需求。简单地说,就是同一内容可以用不同的样式表,显示不同的样式。其技术思路如下:XML(数据描述)+XSL(数据显示格式)+CDF(数据播放周期等)。

然后,可以根据不同用户的不同时间的不同显示要求,编写不同的样式文件和频道文件。这样,每个用户就可按照自己所要求的显示格式和指定的时间间隔获得有关信息。

5 图书馆信息服务频道的技术实现

5.1 频道的生成

一个最基本的完整频道包括:1个CDF文件和3个图形标识文件。第1个图形文件是图标文件ICO,格式为ICO或GIF,用于显示在IE的收藏(Favorites)菜单下的频道子菜单中。第2个图形文件是Image,格式为JPEG或GIF,用于显示在桌面频道条上。第3个图形文件是Wide Image,格式为GIF或JPEG,用于

显示在 IE 浏览器中的频道栏中。如果 Wide Image 不存在,浏览器将使用 Image 来代替。复杂的频道可以包含子频道(Sub-Channel)和项目(Items)。子频道可以让人有层次地组织自己的 Web 页面,针对不同的子频道和项目做不同的设置。

FrontPage 98 提供了一个 Channel Definition Wizard,这样可以很容易地创建一个 CDF 文件。在 FrontPage 内创建一个 Channel Definition Format 文档,步骤如下:

(1) FrontPage Explorer Tools 菜单上,单击 Define Channel。出现 Channel Definition Wizard。

(2) 确定选择了 Create A New Channel Definition Format File For The Current FrontPage Web 选项,然后单击 Next。Channel Definition Wizard 的第 1 步出现,要输入一些信息。

(3) 确认主页的路径显示在 Introduction Page 框内。

(4) 在 Logo Image 框旁边,单击 Browse(浏览)按钮,双击一个 GIF 文件。

(5) 在 Logo Image 框旁边,单击 Browse 按钮,双击另一个 GIF 文件,单击 Next。适当的文件名将写在 CDF 文件中,出现第 2 步。

(6) 确定 Web URL 在 Source Folder 框中列出。出现第 3 步。此步允许操作者从将被送往赞助人的页面的列表中排除页面。

(7) 在相应 HTM 文件上点击,按住 Ctrl,单击另一个 HTM 文件。单击 Exclude,然后单击 Next。出现第 4 步。在此步中,可以修正包括在 CDF 文件中的页面属性。

(8) 确定在 Channel Item 列表中选择了某个 HTM 文件。在 Abstract 框内置插入点,然后键入页面名称。

(9) 确定选择了 Specify Usage 选项和 Channel 控制框,单击 Next。出现第 5 步。在此步中,可以指定数据,向 Web 站点通知赞助人的更新。

(10) 再次单击 Next。出现第 6 步。可以不提供一个 URL 来记录赞助人信息。

(11) 单击 Next,发展到第 7 步。

(12) 在 Additional Option 区域内,确定选择 Place A Button 控制框,并且确定反映新 CDF 文件的 File Name 框被保存在 Web 站点。

(13) 单击 Save。CDF 文件被保存在 Web 站点。赞助人能够看到主页和页面信息,并且使用这种搜索表单发现其他信息。

另外,Microsoft 还提供了 cdfgen 频道编辑工具。

5.2 频道定义文件 CDF

频道最主要的内容是一个频道定义文件 CDF

(Channel Definition Format),该文件是一个文本文件,遵循可扩展标记语言的规范,CDF 文件格式类似 HTML 超文本标注语言。要建立一个频道,只需要用一组 CDF 标签就可以做到。当浏览器在遇到 CDF 文件时,它就自动打开构建向导,帮助用户订购该频道。

6 有待进一步研究的问题

频道技术是一种广播技术,其突出优点是服务的主动性。但它的缺点是缺少个性,也就是服务针对性差。而点播技术正好可以在一定程度上解决缺少个性化的问题,不过服务的主动性也差,每次都要用户自己选择所需的内容。要克服二者的不足,只有在两者结合的基础上融入人工智能、知识发现等技术实现智能化的信息推送。智能信息推送技术的引入,可以提高网络及数据库的智能水平,从根本上解决“推送”和“拉取”技术应用过程中所遇到的上述难题。

在个性化主动信息服务系统中,建立和维护表征用户特点和目标的参数与系统的个性化交互能力有关。这样就存在一个用户模型(User Model)的建立问题。用户建模的目的是为了识别用户的信息(beliefs)、目标(goal)和计划(plan)以提供个性化的服务。目前,通过分析用户以往的需求行为来描述其个性化的需求文件(profile)较为困难,因为用户需求也是一个时间函数,会经常发生变化。将广播技术和点播技术结合起来,更好地主动为用户提供个性化的信息服务,是今后研究的方向。

参考文献

- 李明等. 计算机网络中的 push 技术. 计算机工程, 2000(6)
- 焦玉英, 索传军. 基于“推”模式的网络信息服务及其相关技术研究. 情报学报, 2001(2)
- 彭路, 胡越明. Push 技术及其在 Internet 上的应用. 计算机应用研究, 1999(9)
- 高文等. 数字图书馆——原理与技术实现. 北京: 清华大学出版社, 2000
- Sean Mcgrath 著; 潇湘工作室译. XML 应用实例——建立电子商务应用. 北京: 人民邮电出版社, 2000
- 齐剑锋, 彭岩. 个性化信息服务出利器. 计算机世界, 2001(45)

焦玉英 武汉大学信息管理学院教授, 博士生导师。
通讯地址: 湖北武汉。邮编 430072。

索传军 武汉大学信息管理学院博士生, 郑州大学图书馆副馆长, 副研究员。通讯地址: 河南郑州。邮编 450052。

张学红 南京大学新闻传播学研究生, 河南省电视台编辑。通讯地址: 郑州河南电视台。邮编 450008。

(来稿时间: 2002-03-04)