

刘志敏 董 慧

## 推拉技术在数字图书馆中的应用\*

**摘 要** 推拉技术是一种能提供及时、主动的定制化网上信息服务技术。其原理和工作方法已运用于网上信息服务,在数字图书馆的读者信息服务和信息采访中应用广泛,前景广阔。图1。

参考文献 8。

**关键词** 推拉技术 数字图书馆 信息服务 信息采访

**分类号** G250.76

**ABSTRACT** Pull-push technology is a network technology providing timely, active and customized services. Its principles and methodologies have already been applied in network information services, and can also be applied in reader information services and information acquisition of digital library. 1 fig. 8 refs.

**KEY WORDS** Pull-push technology. Digital library. Information service. Information acquisition.

**CLASS NUMBER** G250.76

推拉技术是一种新的网上信息服务技术,能够提供及时、主动的定制化服务。本文将研究如何将它应用于我们数字图书馆中的信息查询服务、采编工作,使它们更好地服务于读者,同时给我们图书馆工作带来更高的效益。

### 1 推拉技术概述

#### 1.1 推拉技术的概念

所谓“推”技术,是与“拉”技术相对的,就是服务器根据事先规定的设置文件,而不是根据用户实时要求,主动向浏览器递送信息的技术。“推”(Push)

技术与使用浏览器查找的“拉”(Pull)信息技术不同,它是根据用户的需求,有目的地按时将用户感兴趣的信息主动发送到用户的计算机中。就像是广播电台播音,“推”技术主动将最新的新闻和资料推送给客户,使用者不必上网搜索。

在“推”技术问世之前,人们往往利用浏览器在因特网上搜寻,一方面,面对浩如烟海的信息,很多用户花费相当多时间和费用也难以“拉”到自己所需要的信息;另一方面,信息发布者希望将信息及时、主动地发送到感兴趣的用户计算机中,而不是等着用户来拉取。“推”技术采用一种广播的模式,其特

- 
- |   |  |
|---|--|
| <p>13 薛向阳,罗航哉.一种新的颜色相似度定义及其计算方法.计算机学报,1999(9)</p> <p>14,18 刘相滨,邹北骥.一种基于主颜色表的图像检索算法.湖南大学学报(自然科学版),2001(1)</p> <p>15 Jain A K, Vailaya A. Image retrieval using color and shape. Pattern Recognition, 1997,29(8)</p> <p>19 Pass G, Zabih R, Miller J. Comparing images using color coherence vectors. In: Proc of ACM Intern Conf Multimedia. Boston, MA, 1996</p> <p>21 Rickman R, John S. Content based image retrieval using color tuple histograms. In: Proc of SPIE: Storage and Retrieval for Image and Video Database. San Jose, CA, 1996,2670</p> <p>23 Niblack W, Barber R et al. The QBIC project: Querying images by content using color, texture and shape. In:</p> | <p>Proc of SPIE: Storage and Retrieval for Image and Video Database. San Jose, CA, 1994</p> <p>24 Hafner J, Sawhney H, Equitz W et al. Efficient color histogram indexing for quadratic form distance function. IEEE Trans on Pattern Analysis and Machine Intelligence, 1995,17(7)</p> <p>25 Hsu W, Chua T S, Pung H K. An integrated color-spatial approach to content-based image retrieval. In: Proc ACM Multimedia '95 Conference. San Francisco, 1995</p> <p>26 Smith J R, Chang S F. Tools and techniques for color image retrieval. In: Proc of SPIE: Storage and Retrieval for Image and Video Database. San Jose, CA,1996,2670</p> <p>张学福 黑龙江大学信息管理学院副教授,副院长。通讯地址:哈尔滨市。邮编 150080。(来稿时间:2002-07-03)</p> |
|---|--|

\*本文系国家社会科学基金项目“数字图书馆相关关键技术研究”(批准号:00B TQ004)的研究成果。

点是以频道“广播”方式使网上用户得到相同的信息。通常,在网络服务器上有专门的推送软件产品(如 PointCast 公司的 PointCastNetwork),可用来制作欲推送出去的信息内容,并播送出去。在客户端则利用安装在个人电脑中的软件,来接收从网络上来的信息,并显示出来。当有新的信息需要提交时,“推送”软件会以发送 E-mail、播放一个声音、在屏幕上显示一条消息等方式通知用户。使用“推”技术,可以提高用户获取信息的及时性和效率。

网络信息中的“拉”技术是指用户有目的地在网上主动查询信息,用户从浏览器给 Web 发出请求,由 Web 获取所需信息。随着因特网上资源的日趋丰富,面对拥有海量信息的因特网环境,为了能快速、高效地寻找到有用信息,搜索引擎应运而生。搜索引擎是通往浩瀚信息海洋的捷径,因而也有人将其称为“网络之门”。实践表明,搜索引擎是有效的网络信息拉取(查询)的辅助工具。

### 1.2 目前网上推拉技术的应用

推拉技术在网上的应用已经十分广泛,网上信息的查询,商业网站的信息发布等等很多都采用了推或者是拉的技术。现从下面两个方面做一个简单介绍。

#### (1) 推拉技术在电子邮件服务中的应用

目前,一些网站已经运用推拉技术进行电子邮件服务,如著名的 yahoo 网站运用推技术管理邮件,当用户注册 yahoo 的邮箱后,一旦进入 yahoo 的主页,登录后,则会有一个提示框告诉你的邮箱里是否有新邮件,而无需你进入邮箱。

#### (2) 推拉技术在一些商业网站的应用

某些商业网站如“当当”网上书屋经常发一些书籍、光碟等打折广告到我们的电子邮箱里,不管你想不想接受,这些信息都会被发送过来,这就是采用了推技术的信息传播模式。同时,如果你主动去“当当”书屋搜索你所需要的书籍,“当当”将你的查询信息记录下来,根据你的兴趣爱好定期向你发送相关的书籍,或者当你此次查询的书目没有找到,“当当”会在下次有这本书的时候通知你,这都是“当当”为你提供的定制化服务。这是一种用户先“拉”网站后“推”的一种信息服务模式。

### 1.3 传统图书馆中的推和拉

其实,在我们传统的图书馆中也存在着“推拉”,它表现在当读者需要书籍时,会去图书馆根据书籍类别和编目进行查找或由图书管理员代查,这都是“拉”的体现,是用户主动地搜寻信息,图书馆提供的

是一种被动的服务。传统图书馆中的预约体现了推的技术,读者填写好了预约单之后,当他所需要的书回库后,图书管理员就会给他发一个通知单通知他去借书,这是图书馆向用户提供的主动服务。

## 2 推拉技术的原理和工作方式

信息推送技术的主要优点是:及时性好,应用面广,对用户没有技术上的要求;其主要缺点是:信息针对性差,难以满足用户的个性化需求。信息拉取技术的主要优点是:针对性强,能满足用户的个性化需求;其主要缺点是:及时性差,对用户要求较高(需掌握有关的检索技术)。信息推拉技术是将信息推送与拉取两种模式结合,因而能做到取长补短,使二者优势互补。根据推、拉结合顺序及结合方式的差异,又分为以下 4 种不同推拉模式。

### 2.1 “先推后拉”式

先及时地推送最新信息(更新的动态信息),后有针对性地拉取所需的信息。这样,便于用户浏览信息变化的新情况和趋势,从而动态地选取需要深入了解的信息。图书馆将相关方面的新书及其内容简介直接送到读者的计算机,读者获取信息后,登录到该数字图书馆有针对性地搜索相关书籍。这种做法不仅大大提高了数字图书馆的访问率,有利于促进图书馆事业的发展,同时,也有利于我国的文化建设。其工作过程如图 1 所示。

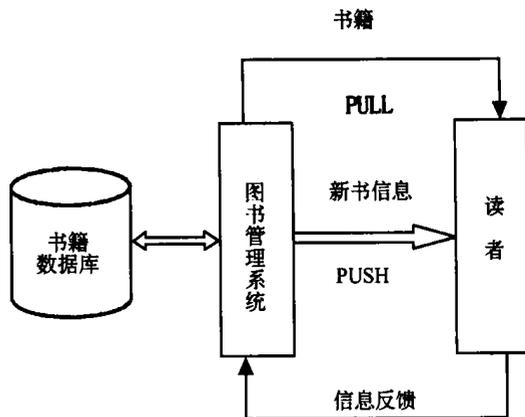


图 1 先推后拉方式在数字图书馆中的应用

### 2.2 “先拉后推”式

用户先拉取搜索所需信息,可根据用户的兴趣,再有针对性地推送相关的其他信息,如个性化的图书订阅。

### 2.3 “推中有拉”式

在信息推送过程中,允许用户随时中断、定格在所感兴趣的网页上,并进一步搜索,主动拉取更丰富的信息。

### 2.4 “拉中有推”式

在用户拉取信息的搜索过程中,根据用户输入的查询信息,信源主动推送相关信息和最新信息。在数字图书馆领域,图书管理员将书籍信息存入书籍数据库,供读者检索查询;将读者需求信息存入需求数据库,对供需信息进行匹配,把经过排序和过滤的信息再推送给用户。信息的匹配是一个复杂的过程,不仅要考虑到书籍的类别,还要考虑读者的文化水平等因素,这就需要利用人工智能技术进行分析和处理,使读者最终得到一组符合条件的信息。

## 3 推拉技术与数字图书馆

“推”技术可将实用的信息“推”给感兴趣的用用户,使用户可以坐等信息到来,它可以实现数字图书馆信息的传播和发布。“拉”技术指用户有目的地在网络上主动查询信息,用户从浏览器给 Web 发出请求,由 Web 获取所需信息。当用户需要某方面的书籍时,只需坐在办公室或家中就可以通过网络进行联机浏览,检索和套录世界上任何一个开放的电子图书馆的信息资料。

因此,推拉技术是我们在数字图书馆有效实现信息和知识检索、传播发布的重要手段。研究更为完善的推拉技术,更方便和智能的信息获取方式是我们数字图书馆中的重要课题。

### 3.1 推拉技术在读者信息获取方面的应用

数字图书馆中的搜索引擎使用户可以直接从网上获取信息,这与传统服务方式相比有较高级别的主动性,但从根本上讲还是被动服务方式。因为,它使用的技术从信息的传输方式看,是由用户通过浏览器服务器发出服务请求,服务器在所拥有的信息资源中进行查询处理,把处理结果传回浏览器所在的计算机来实现的。在“拉”技术中,主动的一方是浏览器,服务器只是根据浏览器的请求,被动进行信息发布。采用了“推”技术的图书管理系统使数字图书馆除了被动地面向整个网络用户服务,还能从技术上主动锁定一批特定用户群,为他们提供专题信息服务,这不但提高了信息服务的效能,还节省了用户在网上漫无边际查询信息的时间。对于一个数字图书馆的 web 站点,只要建设一个专业信息服务频

道,就能够面向自己的用户开展具有很强针对性的主动信息推送服务。一个数字图书馆的专业信息服务频道的工作原理:

(1) 用户登录到图书馆站点,提出获取主动推送信息服务申请(仅第一次获取服务时需要填写)。

(2) 图书馆 webservice 发送一个申请表给用户,具体项目包括:用户名、密码、所需信息的主题、关键词、推送信息的地址、推送周期、要查询的数据库等等。

(3) 用户填写好申请表后,提交给图书馆 webservice。图书馆 webservice 将用户的特征信息、查询要求等传送给“推送服务代理”。

(4) “推送服务代理”根据用户的请求信息,在用户特征信息库(专门记录用户需求特征信息的数据库)和用户信息库(专门记录用户希望获取信息的数据库)中分别增加一条记录。

(5) “推送服务代理”会根据用户的要求,定期将用户的查询需求传递给“查询代理”。

(6) “查询代理”根据“推送服务代理”传送的用户要求,定期检索相应的数据库,并将查询结果返回“推送服务代理”。

(7) “推送服务代理”按照用户的要求,定期将最新信息推送到用户指定的地址。从系统论的角度看,基于 web 站点的信息服务效果的优化是一个系统工程。通过对信息的收集、整理、加工、处理、存储、传递、利用等环节的控制,使信息系统在整体上处于最优状态,都能够改善和优化服务效果。

### 3.2 推拉技术在数字图书馆采访工作中的应用

采访即馆藏文献信息搜集,是图书馆文献信息处理工作的信息输入阶段。采访的工作职能,就是广泛收集各类型文献的出版、发行信息,并根据本馆实际及其读者需求,选择、采集满足读者群现实与潜在需求的文献。传统的文献信息源主要有公开发行的报刊、出版发行单位编印的书目等印刷型书目信息,获得文献的方式主要有预订、邮购、选购等。传统图书馆中,采访人员获得出版单位编印的书目信息是被动的,范围比较小且呈零散状态。在数字图书馆中再采用以前的方式进行采访显然是不适宜的,计算机网络把我们收集文献信息的范围扩大到了全国乃至全世界,除传统的收集方式外,我们还可以通过因特网访问出版发行者的网站,收集文献信息,并在电子商务系统的支持下,实时采集所需文献。这样不但缩短了预订时间周期,而且提高了文献到馆率。同时,我们

还可以利用网上信息推送技术,将我们需要的书籍发布到各出版商的网站,这样,等出版商有了相应的书籍时,就可以及时将书籍信息发送给我们,大大缩短了采集的时间,提高了工作效率。

在图书采访工作中,收集读者需求和意见也是十分重要的。在传统图书馆中,我们很难较好地完成这项工作。因为在传统方式中,收集信息的面不可能太广,只能限于一部分读者,且信息反馈的周期较长,耗费人力大,且收益不太好。如果采用了信息推送技术,我们将征集读者意见做成一个网页,发送到校内每位读者的计算机系统,这样,关心我们图书馆工作,同时也是关心自己切身利益的读者就会给我们回复,这种信息采集、反馈的时间就会大大缩短,所获取的信息也使我们采购工作提供了方向。

#### 4 推拉技术在数字图书馆中的应用前景

信息“推送”与“拉取”两项技术应当取长补短,相互结合。在两者结合的基础上再融入人工智能、知识发现、网络及数据库等技术,从而形成“智能信息推拉”技术。这项技术是当前 Internet/ Extranet/ Intranet、数据库系统及其他信息系统为用户提供信息服务的一个发展方向。智能信息推拉技术的引入,可以提高网络及数据库的智能水平,从而从根本上解决“推送”和“拉取”技术应用过程中所遇到的难题,如如何从海量信息中提取有用信息、如何为不同用户提供个性化信息服务等。

IIPP 技术应用了人工智能(AI)、机器学习(ML)方法、知识工程(KE)的知识推理搜索方法、知识发现(KDD)方法等技术,将“智能信息推送”(IIPush)和“智能信息拉取”(IIPull)相结合。这样,一方面,信息提供者可以推测用户的兴趣,从而更有针对性、更及时地向用户推送实用信息;另一方面,用户也可以更快、更准确地从信源拉取到最新信息。人工智能和KDD技术的应用,提高了网络的智能化水平,可以在“推送”或“拉取”到的大量信息中发现其内在规律,提取用户最关心、最感兴趣的有用信息,从而大大降低用户进行下一步搜索和筛选的工作量,进而提高用户获取信息的效率和能力。

数字图书馆未来的发展趋势总的来讲是图书馆藏品(文字、图像、音频和视频等)的存储和组织的数字化、搜索和利用的网络化,从研究内容来看,当前的数字图书馆研究侧重如下三点:

首先是数字化藏品的加工处理的自动化、智能化。当前图书馆藏品和因特网资源的分析、摘要和索引主要是由手工完成的,工作繁重。对它们自动化的聚类分析、以内容为基础的自然语言分析、摘要和索引尚处在初级阶段,这方面的研究将大大提高图书馆自身的工作效率甚至工作质量。

其次,与席卷全球的变革浪潮相对应,数字图书馆的研究应该强调创新。数字图书馆存在的意义不仅是传统图书馆服务媒体和方式的转变,它还利用自身数字化和网络化的优势,提供新的思路和工具,增加数字图书馆的功能,研究为教育、学术研究进行服务的新途径,如电子期刊、电子书籍和以虚拟教室、实验室、博物馆、音乐厅、画廊等形式出现的在线教育和文化源、跨语言和全球性的知识库等。

第三,以人为本,极大地关注人在应用数字图书馆时的感受,处处为用户着想,研究人机关系、人机界面、强调个性化服务,减少在利用数字图书馆过程中用户的干预等。

信息服务包括的功能如导航、过滤(即从大量信息中筛选符合条件的信息)、整理(即为用户将资源进行分门别类的组织)和知识发现。这些都是使信息服务走向个性化主动服务不可缺少的功能。

采用“Push + Pull”的方式,服务器不仅要把信息推给客户,而且还能够按照客户预先设定的触发事件和发送要求,在条件满足时自动向客户发送消息。真正的个性化应该是动态而主动的,在最初的规则制定好之后,系统能够自动跟踪用户的使用倾向。结合人工智能的推拉技术能够让我们图书馆为读者提供更好的个性化服务。

#### 参考文献

- 1 黄秀霞. 网络环境下的数字图书馆建设. 情报科学, 2001(1)
- 2 段立娟, 刘桂林. 数字图书馆. 中国图像图形学报, 2001(8)
- 3 索传军, 焦玉英. 基于 web 的信息服务效果的优化研究. 中国图书馆学报, 2001(7)
- 4 张自生. 信息系统及其网络环境下的图书馆编目工作. 蒙自师范高等专科学校学报, 2000(2)
- 5 涂海清. 网络环境下高校图书馆编目工作的特点及对策. 常熟高专学报, 2001(1)
- 6 彭岩, 艾迪明. 网络信息“拉取”技术. <http://www.ccw.com.cn/>
- 7 彭岩, 鄢琦. 网络信息“推送”技术. <http://www.ccw.com.cn/>
- 8 涂序彦, 曹斌, 陈鸿绢. 智能信息“推一拉”方法与技术. <http://www.ccw.com.cn/>

刘志敏 武汉大学信息管理学院硕士生. 通讯地址: 湖北武汉. 邮编 430072.

董 慧 武汉大学信息管理学院教授, 博士生导师. 通讯地址同上. (来稿时间: 2002-07-26)