

信息推送技术及其应用

沈艺 (南京师范大学 210097)

摘要:本文分析了 Internet 上信息推送技术(push)的运行机制,以及不同的实现形式,最后讨论了目前信息推送技术存在的问题和今后的对策。

关键词:信息推送 Internet 情报检索

众所周知,Internet 是我们获取信息的有效途径。通过 Internet 可以搜索任何一台上网服务器,获取信息。Internet 发展迅速,网上信息极度膨胀,查找信息犹如大海捞针。虽然 Yahoo、Infoseek 等提供了优秀的搜索引擎和大量的信息资源,但仍需根据检索到的地址信息,按照屏幕页面的提示一步步地向下追踪,如果查到的地址有几百条、几千条甚至更多,要确定哪个地址符合要求就很困难。

因此,人们迫切希望找到一种能够在信息海洋中自动获取实用、准确、精炼和优质信息的方法。信息推送技术(push)正是面对这样的需求而产生的,它将深刻地改变互联网络的传播方式,并对信息采集、加工、传递、利用和控制产生重大影响。

1. 运行机制

所谓信息推送,就是网络公司用一定的技术标准或协议,从网上的信息源或信息制作商获取信息,通过固定频道向用户发送信息的新型信息传播系统。应用信息推送技术可以建立网络广播站(Webcasting),即网络信息广播系统,它通过智能化的代理服务器从 Internet 上不断取回用户所需的信息,将信息进行分类,在主机上设立了固定的“信息频道”和“信息树”,供用户对网上的信息进行预定和选择,用户联网后,可随时获得经过更新的各类信息。

信息推送技术正在改变 Internet 上信息访问的方式,它变用户主动搜寻信息为被动的、有目的接收信息。推送技术不仅是一种单纯的信息提交技术,它还可以把音频、视频、Web 服务器中的信息以及数据库中的数据捆绑在一起,在防火墙内外向用户提交丰富的多媒体信息。为此,“push”客户机软件允许用户定制从各种信息源向其提交的信息。

目前多数“push”客户机软件可以向用户提供最新的新闻信息,而这一切功能是以以前任何浏览程序所无法完成的。“push”客户机软件可以对信息进行分类,并可实

时地向用户发布最新的信息。客户机操作过程是:凡当前新的信息内容可获得时,客户机便可自动被告之。大部分客户机软件采用预约模式,即这些客户机软件均按预定义的时间间隔定时询问提供信息的服务器,以查询当前是否有新的信息内容能够加以提供。

用户还可以对这些客户机软件的接口进行定制化处理,以使它们所提交的信道成为专用信道,即令每一信道只传送某一信息内容提供商所提供的某一类信息。当新的信息需求被提交时,“push”客户机软件将采取以下方式通知用户:通过 email、播放一个声音、显示一个图符弹出某一应用、激活一个屏幕保护程序、显示墙纸或通知单等。

2. 信息推送的形式

推送技术的最基本的形式是一个简单的通知(notification),如电子邮件。针对这种服务,用户可控制它通知的形式、时间间隔等。通知并不具备很强的交互性和强制性,对资源和信息流量的要求不高。

比简单的通知智能化程度更高的推送技术是一种提要。提要可实现查看 Web 页或其他信息源,寻找需要匹配的信息、并向用户传递信息。用户要以关键词、日期、数值、比较规则以及其他查询条件提供要查找信息。提要有很多后台进行的处理活动,不仅是给用户每天一次的报道,它的处理活动还要受查找条件的制约,这些后台处理过程与用户的联系是不可预测的。

推送技术的第三种形式是自动拉出。它有一组可供用户经常查看的 Web 页。自动拉出将获得所有这些 Web 页,并保存起来供用户以后阅读。自动拉出可以获得许多材料,用户还可以通过 email 接收这些材料,或至少通过 email 知道这些 Web 页是为你编制的。

自动推送是一种高级推送技术。自动推送能够根据自身的刷新时间表发布信息。用户可以预订推送信息服务。但需要在 Web 上连续收听广播。在一般情况下,这种服务要求在用户终端上装有特殊的客户机软件,定期

发出更新请求。如果用户不在 Web 上提出服务要求,他将得不到任何服务。从自动推送,用户得到的可能是全屏报道,或在屏幕底部显示大字标题。这种级别的推送技术有很多交互性,用户可以选择需要查看的信息流,也可以精选发送给用户的信息,或者试探可能让用户感兴趣的其他信息。

在这几种形式的推送技术中,自动拉出常需要在本机上保存所获得的 Web 页,这就要求用户在设置要查看内容和复查收到的 Web 页方面进行更多的工作和交互活动。自动拉出与提要不同,因为实际的 Web 页会不断显示出来供用户逐一浏览。此外,自动拉出还要求更多的资源来保存信息。

通知会每天准时告诉用户信息。提要使用的媒体与通知类似,但提要具有更强的指导意义,因为用户可以查看具有确定时效的信息。提要在设定条件时会给用户更大的交互性,而对资源的影响很小。用户可以通过筛选安全的 Web 信息,建立相应的服务,例如:Individual 的 Newspages(预订)、Excite Live 的 NewsTracker(免费)、MIT Media Lab 的 Fish, Wrap(免费)、Netscape 的 In-Box Direct(免费)以及 Yahoo! 的 My Yahoo(免费)等。

3. 频道按钮与频道定义格式

高级的信息推送采用的是一种频道转换技术。用户只需要改变频道就可以选择气象、体育、金融或其他信息。用户浏览器必须向远程服务器发出“update me”信息流。频道转换技术的最大采用者是 Microsoft 公司和 Netscape 公司,它们都率先采用这项技术。其他公司也试图在这一技术中占有一席之地。

可以预言,频道转换将是 Web 上最流行的推送技术,而这种技术将有两种标准,即 Microsoft 的标准和 Netscape 的标准。采用这两种标准提供的所有信息内容都可以被任何用户使用。

Microsoft 已经推出自己的频道管理“标准”,即频道定义格式(CDF)。CDF 包含为部门或团体创建频道的能力。CDF 能自动地把原有浏览器书签转换成频道,使得用户过去找到的有用站点,机构中的其他人也可以使用。

不管 CDF 是否被定为这一领域的标准,Netscape 正在这一领域积极推广它的 Java/HTML/HTTP 方法,从而使这一领域可能会有两种标准。

Netscape 的推送技术是以 Netcaster 为基础演变而成的。Netcaster 支持的频道模式包括 Dynamic HTML,并具有分层、字体变化和画布方式。其 Channel Finder 能够让用户查看和预订现有标牌频道,由服务管理人员掌握通向标牌频道的钥匙,他们可以制作可用通道,并限定对

内部专用信息的存取。Netcaster 能够执行自动 pull,获取和保存任何一个站点或通道,供以后脱机时查看。

4. 简单信息推送的实现

根据计算机网络条件及用户分布情况,可建立信息推送服务实验系统。下面以书目信息的推送服务,简单说明实现过程。

首先开发用户注册子系统和邮件定时自动推送子系统。根据用户注册的情况,对中心服务器中的书目信息,按用户的要求进行手工检索分类,以文件形式存档,系统自动将存档文件推送到用户的 email 信箱中。

在手工分检的基础上实现书目信息的自动分检。也就是说根据用户发送的查询条件自动将书目信息检出,存在相应的文件中,由系统自动推送到用户信箱。信息自动分类极大地减轻了劳动强度,提高了信息的时效性。

今后将建立基于 Web 的信息推送,它的界面更友好,功能更强。

5. 问题与对策

自动推送有时会占用太多的本地存储资源,而且会对信息流量产生不良影响。连续地搭接 Web 和有规律地传送 HTML 文件,这是 Web 接入存在的主要问题。如果在一个机构中有许多人都在从事同样的工作,问题就更严重,因为所有资料都要占用网络和磁盘资源。

同样,自动拉出也要提供 Web 页的不断显示,来供用户逐一浏览,这要求更多的空间来保存信息。如果有许多人都在进行这项活动,也会影响信息流量。

因此,在实际应用中,要根据不同的情况改变信息更新时间和频率。也可以将系统设置成自动更新状态,使更新在机器闲置时进行,这一功能可以有效地交错接入 Web,而防止信息阻塞。如果要查看信息的时间性不是很强的话,信息刷新就可以隔夜进行。用户只需要让机器处于打开状态,并对软件进行配置,告诉它何时刷新信息,这样就可错开高峰期接入 Web。

此外,Internet 上的信息分类、过滤是实现信息推送的关键,而 Internet 上中文信息的分类、过滤则是目前计算机领域的研究重点,目前还没有相对成熟的技术及软件,这需要各方面共同协作,以推动我国信息技术的发展。

参考文献

- [1] 郭力等, Internet 网络系统与资源, 北京: 科学出版社, 1996
- [2] Johnson 等, Visual J++ 开发人员指南, 北京: 机械工业出版社, 1997

(来稿时间: 1998 年 12 月)