

智能Web 站点技术研究

利玄英, 李名世

(厦门大学计算机科学系, 福建 厦门 361005)

摘要: 为了使Web 站点进一步适应当前复杂的电子商务, 必须引入智能技术。通过融合XML 信息交换、智能文档分析与搜索、数据库动态网页、数据仓库、虚拟现实和流式多媒体等技术, 将能够建立起支持生动的个性化服务、智能搜索与信息推送、实时交互与协同工作、事务处理与决策支持、通用数据访问与数据挖掘等功能的智能站点。

关键词: Web 智能站点; 电子商务; 个性化服务; XML

中图分类号: TP391

文献标识码: A

1 传统HTML 的不足

WWW (World Wide Web) 是在1989- 1991 年间由在CERN 的Tim Berners 为物理学家们的协同工作而开发的一个信息系统, 在NCSA Mosaic 客户机被推出后迅速发展起来。由于超越了常规应用程序所面对的硬件和操作系统依赖问题, 它在Internet 上的广泛传播导致HTML 浏览器成为应用的一个新的传送平台。WWW 的结构是基于超文本的Client/Server 间的信息传递。超文本用HTML 编码, 客户机与服务器使用HTTP 的轻量协议进行交流, 内容寻址则使用URL 标志法。

HTML 继承了SGML (标准通用标记语言) 的一些重要优点。作为一种相对简单的标识语言, 它为WWW 的迅速发展作出了贡献。但HTML 有诸多不足, 主要体现在如下几个方面: 1) 固定的文档类型, 扩展性差。不允许用户设定自己文件的标签或者属性, 因而难以表达各种抽象内容和提供个性化服务; 2) 超链接(即在任意的信息元素间创建并传递联系) 功能较弱, 不支持描述数据库和面向对象层次的深层结构规范; 3) 搜索程序常需理解各种数据库的模式, 难以进行一些有意义的检索; 4) 数据确认方面不支持检查输入数据合法性的语法规范。

传统HTML 过分限制了Web 文件的复杂性和灵活性, 使之难以满足商务上的客观需求。为了克服其不足, 人们已经开发出大量可扩展HTML 语言功能的解决方案, 但因依赖于不兼容的专利技术或格式, 通常难以进行维护, 而且还有可能使性能降低。电子商务的发展迫切需要在技术升级换代的智能型网站的出现。

收稿日期: 2000-06-21

作者简介: 利玄英(1977-), 女, 硕士研究生

2 智能站点的主要特征

区别于传统Web站点,以下提出智能型Web站点应具有的主要特征:

1) 用户鉴别与个性化服务

千篇一律的商务站点缺乏传统商业活动中人的情感交流和互动,就如大量的没有针对性的印刷广告,难以吸引顾客而被当作垃圾邮件一样。有针对性地开展个性化服务,才能使商务站点吸引更多的用户。个性化站点对用户的服务包括:a)个性化操作与显示,用户可以根据其个人爱好,选定特定的操作与显示方式。新一代的Web文档中,显示格式不再是内嵌在文档数据内部,而是存放在与文档的数据相分离的样式文件中,文档可根据不同的样式文件,因人而异地控制显示方式,使显示具有多样性和灵活性;b)个性化问询与导航,每当用户登录,网站会根据各个用户的特定情况给出有针对性的问询与导航;c)个性化信息推送,网站根据用户自定义或自动分析用户最可能需要的信息,在用户预订或进行一定操作后,自动把信息推送给用户。

个性化服务的优点是明显的:首先,对用户的服务质量大大提高,个性化商务网站将包含所有注册用户信息的数据库。针对不同的用户,可以采用对该类用户最优的用户界面与交互方式,并展示不同的信息与广告,使之到达最需要该信息的用户手里,避免大量无用信息对用户的干扰。其次,网络信息流量减小,响应速度大大提高。传统的万维网模型中,用户使用象IE或Netscape这样的浏览器直接与远程服务器上的数据源进行交互,在整个会话过程中,客户程序必须维持与服务器的连接。对每个查询请求,服务器都发回一个HTML形式的回答供客户浏览器直接浏览。但在个性化服务中,由于服务器了解用户的个人信息,因此可以建立三层模型。中间层软件一次性地从远程数据库中获取满足客户需要的信息,接下的会话过程中间层软件就不需要与远程数据库进行交互,因而可以中止与远程服务器的连接。中间层软件则把满足客户需要的信息组装并一次发送到客户端,然后与客户端断开连接。客户端应用程序直接使用中间层软件发送的组装信息,对用户的操作提供基于客户端的响应,而不再请求服务器或中间层软件,这样对用户的响应速度大大提高,并极大地减少了网络流量。

2) 智能搜索引擎与信息推送活动频道

智能搜索引擎能够自动收集和记录用户的兴趣主题,根据不同的用户给出不同的导航建议。当用户有信息需求时,可以紧扣用户的需求,提高用户的检索速度和查准率。在这过程中,它把检索任务细化,并综合运用各种智能化专业检索引擎,如医学领域中查找某类疾病,电子商务中查找最低价格图书的检索引擎,根据不同用户对各类检索引擎赋予不同的权值,并对检索的结果进行分析和综合。

智能搜索引擎提供快速响应用户信息请求的方法,而信息推送活动频道则使自动传递用户定向信息成为可能。一旦用户订阅了某一频道,该频道的内容变更将定向地推送给用户。

信息推送活动频道应具有以下功能:a)通知频道自动通知用户Web站点内容的变更;b)专用频道自动下载Web站点内容供用户脱机浏览;c)活动桌面项目提供追踪和观看分类信息的功能;d)根据用户预定的感兴趣的信息,为用户创建个性化频道或活动桌面项目,将其感兴趣的信息推送给用户。

3) 开放式超文档与超链接结构

传统的Web文档使用超链,通过定位点,可以从源文档转到目标文档。目标文档则无从得

知这个指向它的源文档的存在。新一代智能站点使用开放式超文档,通过扩展超链接来关联。扩展超链接可以存放在第三方,使源与目标两端都感知链接的存在。通过扩展超链接,可在不改变Web 文档本身的情况下给它加上标注,并可共享。使用开放式超文档,Web 可以增加以下优点:首先,资源的交流更为灵活。对Web 的批注文档可以成为其他用户遍历Web 的指南手册,可为其他用户提供一些找到更新材料的建议,甚至控制浏览器的链接遍历能力,以阻止不良信息传播。其次,文档管理更具智能性。如使用扩展超链接把Web 资源与其属主的有关许可政策的元数据联系起来,当资源被卖出或被许可使用时,元数据可能会发生变化,但资源本身不必作任何改动。以此,可大大减少对资源的非故意破坏,并使属主和管理政策的改动更为简单。

4) 实时交互与协同工作

传统的Web 交互中,Web 并不支持持久的连接,每次用户请求文档或数据文件时,服务器发送响应但不记得用户曾访问过,从而无法标识用户并同用户对话。这种特性使得站点对多用户在网上传协同工作的支持十分有限。智能站点应采用各种途径在Web 上建立用户上下文,强化对CSCW 技术的支持,使得网上实时的商贸洽谈、协同著作、协同设计可以非常便捷地进行。

5) 支持通用数据访问与数据挖掘

智能Web 站点不仅要有用户端智能,也应具有业主端智能。它不仅是站点发布方与最终用户的交互界面,而且是发布方进行内部管理和决策论证的强有力工具。因此,站点除了集成实时运作的数据库系统外,还必须提供对数据仓库和数据挖掘的支持。由于数据的大量堆砌将影响事务处理的效能,故应由数据仓库来存储和归纳历史数据,以提供深层次的数据挖掘和决策支持。

6) 融合虚拟现实和流式多媒体技术

虚拟现实利用计算机构成一种身临其境的模拟环境,通过多种传感器技术实现用户与虚拟环境的自然交互。虚拟现实有助于建立和谐的人机交互环境,使人们的认识空间与计算机的处理空间尽可能一致。智能网站对虚拟现实的支持将极大增加网站的亲和力。随着网络带宽的提高,网站还应引入流式多媒体技术,提供比静态的文字和图象更具说服力的视音频信息,更好地阐述某些概念或过程。

3 基于XML 的信息交换

作为一种相对简单的标识语言,HTML 的局限性使得仅靠它难以构造功能复杂的智能站点。最近推出的可扩展标记语言XML,为Web 技术的发展注入强大的生命力。XML 的如下特点使得它可望成为未来Web 开发的主流语言:1)良好的可扩展性,XML 是一种元标记语言,它定义了一组用来形成语义标记的规则集,可以用来构造特殊领域的标记语言;2)高度的结构化,通过文档类型定义DTD 或XML Schema,XML 不仅可以指定文档中的元素,还可以指定各元素之间的关系,这种结构化对大型数据库应用特别有用;3)强大的超链接,XML 的链接语言XLL 支持可扩展的链接和多方向的链接。能支持独立于地址的域名、双向链路、环路、多个源的集合链接等;4)文档的表示形式多样化,XML 将数据本身与数据的表示相分离。数据表示是由层叠样式表CSS 或可扩展样式语言XSL 来实现的,可使网络的用户界面更趋于个性化、风格化;5)将大量运算负荷分布在客户端,使得服务端可以尽可能完善、准确地将数据封

装进 XML 文件,并使广泛、通用的分布式计算成为可能

将 XML 嵌入乃至取代 HTML,智能站点中的个性化服务、双向链接、快速搜索引擎以及在线数据分析处理(OLAP)功能可望获得较好地实现

4 结 论

复杂的商务站点不应沿用综合信息网站的传统技术来构造。集成了众多最新技术的智能 Web 站点将对新型 EDI B to B (企业到企业)供应链集成、比较购物与发现代理等众多热点问题提供解决方案

参考文献:

- [1] Charles F. Goldfarb, Paul Prescod, The XML Handbook[M]. Prentice Hall PTR, 1998
- [2] 姜旭平. 网络商务处理系统[M]. 北京: 人民邮电出版社, 1999
- [3] 王珊. 数据仓库技术与联机分析处理[M]. 北京: 科学出版社, 1998
- [4] Kalakota R, Whinston A B. Frontiers of Electronic commerce[J]. Addison-Wesley Publishing Company, 1996

Study on the Technologies of Intelligent Web Site

L I Xuan-ying, L I M ing-shi

(Dept. of Computer Science, Xiamen Univ., Xiamen 361005, China)

Abstract: In order to make the Web site more suitable for the increasing complex tasks of E-commerce, the Web site should be more intelligent than it is now. By integrating the technologies of Information exchange based on XML, intelligent documents analysis and search engine, dynamic web pages, Data Warehouse, virtual reality and stream media, we can build up an powerful intelligent web site that supports personalized service, intelligent search engine and information push, real-time interaction and cooperation, transaction processing and decision supporting

Key words: intelligent Web sites; electronic commerce; personalized service; XML