

基于 ISAPI 的社会科学引文索引短信平台服务

朱 超 苏新宁 邓三鸿

(南京大学信息管理系 南京 210093)

【摘要】 社会科学引文索引系统(CSSCI)是社会科学科研水平和综合实力评价的重要方式之一,而短信平台的拓展既是服务方式的扩充也是考虑用户的需求而致。本文通过 ISAPI 和普通 B/S 模式的一系列对比,分析为什么选择 ISAPI,并详细论述了系统的实现,展示如何向用户提供信息查询、推介等服务。

【关键词】 CSSCI ISAPI 短信平台 信息推介 **【分类号】** G354

Service by CSSCI Short - Message Platform Based on ISAPI

Zhu Chao Su Xinning Deng Sanhong

(Information Management Department, Nanjing University, Nanjing 210093, China)

【Abstract】 CSSCI System is regarded as one of the important methods on evaluating research level and integrated strength of social science. It has received solid approbation after appearance with its influence amplifying day by day. Meanwhile, short - message platform is founded considering aggrandizing service channels and meeting users' needs. This paper analyzes the reason why we choose ISAPI as solution by the comparison between ISAPI and common B/S schema, and finally gives the realization and explains how to provide service including information retrieval and dispatch to users.

【Keywords】 CSSCI ISAPI Short - Message platform Information dispatch

1 概 述

中文社会科学引文索引(CSSCI)自创立至今,经过不断的发展完善,目前已经可以通过多种途径向用户提供服务,如网络检索、光盘检索、定题服务、参考咨询等。随着社会信息化程度的加深和人们对信息时效性要求的提高,作为移动信息服务手段之一的短信也有了进一步的发展。由于短信息具有快速、便捷、双向性的特点,使其在各个领域得到了迅速的发展^[6]。同时,与短信息服务相比,定题、参考咨询这些服务都属于异步推送,即信息发送和用户接收不同时,而且由于受计算机和网络环境的限制,常常使信息的时效性大打折扣。因此,在某些实时性要求较高的情况下,如同步推送服务的实现,短信服务成为首选。于是如何利用短信息为用户提供优质高效和多样化的服务,拓宽服务渠道,并构建相应的短信息平台成为我们关注的课题。

2 选择 ISAPI 作为解决方案

2.1 B/S 模式及其缺陷

目前基于 WWW 服务的应用越来越多。这些应用系统的结构也由传统的 C/S 模式发展为三层结构的 B/S 模式。B/S 模式采用浏览器/服务器结构的网络计算模式,把所有的开发与计算工作转移到了服务器端,从而达到“瘦客户端”的目的。但事实上,仔细分析就会发现该模式存在安全上隐患,主要表现在这逻辑事务层就要求我们采用有效的方法和措施消除安全隐患。

在 Windows NT 环境下,逻辑中间层是由 IIS 提供的。IIS 可以提供基于 WWW 的服务。同时可以执行的多种脚本满足用户动态内容的交互。由于这一层是处于服务器端的,就很可能造成安全漏洞问题。目前 IIS 出现过以下典型的安全漏洞:存在一种被称为“非法 HTR 请求”的缺陷,此缺陷在特定情况下会导致任意代码都可以在服务器端运行;其次,通过构造特殊字符,可以查看 ASP 源代码。IIS5 之前版本的一个广为人知的漏洞就是 \$DATA。通过它使用 IE 的源代码查看或 Netscape 直接访问该 ASP 文件就能轻而易举地看到 ASP 代码。

数据源层一般由 SQL Server 或 Oracle 数据库来负责数据的存储工作。在这一层也同样存在很多的漏洞。比较常见的是 SQL Server 的 1433 端口远程连接安全漏洞。一旦知道了数据库用户的口令或者系统用户的口令就可以通过远程连接来存取远程的数据库数据。

由此,我们想到兼具 B/S 特性,又同时采用代码封装和数据过滤功能的 ISAPI 能很好地处理这些问题,是我们的很好的解决方案。

2.2 选择 ISAPI 改进 B/S 模式

ISAPI 是由微软和 Process 软件公司联合提出的 Web 服务器上的 API 标准,它与 Web 服务器 Internet Information Server(IIS)结合紧密,功能强大,利用 ISAPI 可以开发出灵活高效的 Web 服务器增强程序。ISAPI 主要分为 ISAPI Server 和 ISAPI Filter 两部分。ISAPI Server 以类似 CGI 的方式,指向服务器的作业,供程序开发人员设计一些扩展功能;而 ISAPI Filter 则倾向于构造服务器。总结 ISAPI 程序的主要特点就是以 DLL 的形式被 Web 服务器加载到进程中,与服务器结合紧密,共用同一个地址空间,共享同一个进程上下文,主要通过一个参数块与服务器进行交互,可以从服务器那里获得关于当前 HTTP 连接的大量信息,在 Web 服务器处理用户请求前就截取它,能够对服务器和客户端之间的通信进行预处理和后处理。灵活定制这些处理,可以实现许多依赖防火墙才能实现的功能,定制一个自己的信息过滤“防火墙”,为本地或远程用户提供经过认证与授权的对网络资源的访问^[4]。

比之普通的 B/S 模式提供的服务,这种工作模式有以下显著的优点:

(1)在整个模型中,增加了 ISAPI Filter 过滤层,形成四层网络结构。通过自己定制相应的处理实现过滤和保护系统的目的。ISAPI 过滤器可以定制的处理有接收 HTTP 协议头预处理、发送 HTTP 协议头预处理、获得原始数据预处理、自定义的安全认证机制、对通信进行加密/解密、自定义日志记录等^[3]。其运行原理决定了它能实现原先脚本预言很难实现的许多保护和自定义功能。

(2)主要的业务处理逻辑都以 DLL 的形式封装起来,从加载时候起一直驻留在内存中,对用户来说完全透明,很大程度上能避免代码泄漏情况的发生。而且,由于 ISAPI 采用的是线程访问机制,一个用户请求就是一个线程,从而能够实现系统资源和动态链接库代码共享,减少了进程调度和进程间通信(如 CGI)等诸多问题,降低系统开销,大大提高执行效率。

除此之外,我们将数据库端保护起来,只有 ISAPI Server 可以访问,任何对数据库的访问或更改操作都只能

通过 ISAPI Server 才能进行,对于每一次访问和更新,都将请求和操作记录到日志中,从而达到有效实现保护数据库的目的。

3 短信平台的系统实现

3.1 系统配置

向当地电信运营商(如中国移动、联通)申请专用网关,并可自行定义消息端口,它将负责用户短信的收发过程。系统服务器内存要求至少不低于 512M,内存过小将严重影响运行速度,本系统中的 SUN 服务器内存为 2G;其他硬件采用中档以上服务器配置即可。

服务器软件配置为 Win2000 Advanced Server + IIS5。

CSSCI 数据量庞大,前期数据加工整理繁琐,因而我们有针对性地采用北京快威数码公司的 DIPS 系统作为后台数据库,较好地完成数据加工入库的工作,简化了前期处理和后期利用。实际平台建设中,若不涉及这些问题完全可以选择市面上流行的 SQL Server、Oracle 等数据库。

开发工具,主要使用 VS.net。

3.2 短信平台系统流程及模块设计

我们给出系统的简要运行流程图(如图1):手机用户发送指定字符串到其所属短信息网关,网关端口接受信息后选择对应的服务提供者并发送查询信息,服务提供者通过请求代理模块解析查询需求,将其转换成对应的 HTTP 请求发送给网络服务器,网络服务器从请求的 URL 中获得参数并从数据库中获取最终结果,最终沿来时途径返回,到达手机用户中止。在此过程中,网络服务器功能实现是最主要的模块,它完成处理请求和数据获取的任务,是研究的重点。

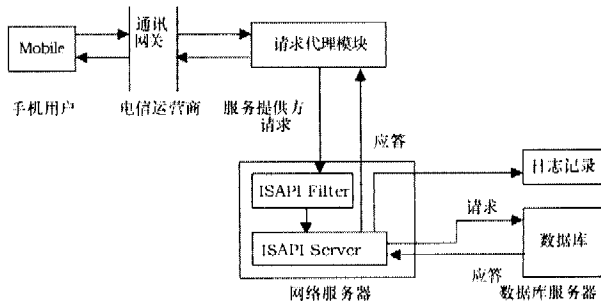


图1 系统运行流程图

(1) 短信网关系统组成

信息服务平台与手机间信息传送随着手机广泛使用,网络技术的发展成熟以及通讯费用逐渐降低而发展迅速。这种服务又可以通过两种途径来实现:一是通过网关收发短消息,该方法不需要附加的硬件,但服务提供商需要到电信部门申

请网关;二是通过 GSM Modem 收发短消息^[5]。本系统正是建立在第一种途径基础上而设计实现的,系统中共有四种提供服务的角色:

服务提供商(SP),它是短信息服务的信息提供者;

短信网关(ISMG),它主要提供 SP 与短信中心之间数据交换的通道,采用 CMPP2.0 协议(以移动通信为例)与 SP 进行连接,采用 SMPP3.3 协议与短信中心连接;

短信中心(SMSC)利用信令网将短信发送给手机用户;

此外还有汇接网关(GNS),主要提供短信网关的路由查询。

由这四种角色支撑的短信息服务的业务过程共有两个部分:

用户发短信业务—MO 业务,短信中心将短信内容以 SMPP 协议发给连接的短信网关。短信网关根据短信发送号码(即 SP 的服务代码)将短信内容以 CMPP 协议发给相应的 SP,由 SP 进行处理。

用户收短信业务—MT 业务,提供此项业务的 SP 将短信内容以 CMPP 协议发给连接的短信网关,由短信网关根据用户手机号码以 SMPP 协议发给用户归属的短信中心。

需要说明的是,如果用户归属的短信中心和提供业务的 SP 连接在不同的短信网关上,则由短信网关之间以 CMPP 协议进行转发。

我们可以用图2清楚地表示网关系统的逻辑结构和运作流程。

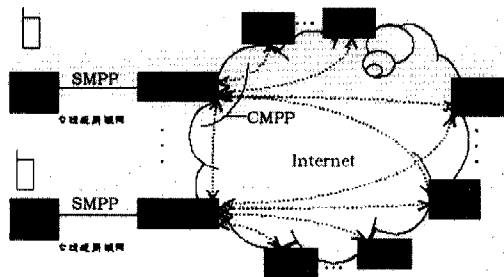


图2 互联网短信网关逻辑结构

(2) 请求代理模块

请求代理模块负责通讯网关和服务平台的衔接,所有的请求应答都通过它来完成,是系统中最重要的部分。其中,ISMG、SMSC 和 SP 之间都是采用 SOCKET 套接字进行通信,用的是常见的客户机/服务器方式。通信前创建了各自的套接字建立连接,启用独立线程侦听外部请求,然后对相应的套接字进行“读”和“写”操作,实现信息的交换。在这里短信网关互为客户服务端。当 SMSC 和 ISMG 通信时,遵守的是基于 TCP/IP 应用层协议—SMPP 协议,网关作为客户端主动和短信中心连接;当短信网关和应用业务提供商之间通信时,遵守的是基于网关相连。他们之间通信采用长连接,即在一个连接上可连续发送多个数据包,然后再断开连接。

当用户发送查询信息时,短信平台检测到请求信号,就由平台路由到请求代理模块,请求代理模块查询数据库的数据表,找到用以区分用户的查询请求的业务码所对应的 URL,然后发起 HTTP 请求,必要时也可能调用系统所提供的 HTTP 下发接口(WEBSend,SENDICP)。在本系统中,因为只提供一种查询 URL,所以请求代理模块直接调用 HTTP 服务,发送数据库查询请求,完成查询后将返回信息送到短信业务平台,短信业务平台再将其送到上一级短信网关。

3.3 短信平台系统实现

手机用户主要是利用 CSSCI 能够为人们提供研究成果的渊源(通过来源文献获得引用文献)、某一科研成果的社会影响和效应(由该成果的被引用情况分析得到)、某个学者的社会影响(该作者发文和被引用的情况统计)、某学科课题研究进展等诸多方面的信息^[1]。故我们也从不同途径提供统计结果:作者总计发文信息,并分年统计;作者被引文统计信息,并分年统计;作者发文任一篇信息;作者被引文任一篇信息并可继而得到来源文献信息;既定主题词查找;某一年某学科学者发文排名,全文索取途径等等。随着需求的增大,可以很方便地逐渐扩充其功能。

具体实现过程中,基于数据安全性的考虑,我们将所有 ISAPI Server 对数据库的访问行为封装成静态库的形式,只提供各种数据访问的接口,包括字段级、记录级、全文级以及多媒体等访问接口,并对所有数据设定访问权限,只有被授权的数据才能允许访问,其权限控制在 ISAPI Server 中实现。同时,将用户手机号码,查询内容,访问时间、成功与否以及异常信息等详细地记录到用户访问日志中,跟踪用户行为。

举例说明:

规定在查询作者苏新宁发文信息总计时,用户只要发送字符串“LC 苏新宁”到 95590(可在申请网关时确定)即可获得该作者发文统计信息,而不需要知道其中是怎样实现的。

通讯网关接收到此命令,立即发送给能处理此命令的服务提供者,服务提供者在请求代理模块中将此命令转化成 HTTP 请求,即

“http://cssci.nju.edu.cn/ShortMsgRetrieval.dll? PhoneNO=13 * * * * * &Qstring=LC 苏新宁”,ISAPI Filter 对请求进行安全性过滤,通过检查后,在 ISAPI Server 中,DLL 内部截取此 URL 中的参数,即 PhoneNO 和 Qstring,利用静态库中的访问接口访问数据库并返回信息,成功则然,否则给出不成功的提示信息,让用户决定下一步的程序。在 ShortMsgRetrieval.dll 中有 ShortMsgRetrieval 函数,此函数是该 DLL 的默认执行函数,它负责识别所有的查询要求并调用对应函数,本例中它先识别出手机号码和查询种类“LC”,随即

调用作者发文统计函数,便得到作者历年发文总数及每年发文数,最后将此结果返回。本例结果形式为:苏新宁,发文计**篇,1998年*,1999年*,...联系电话:*****。

又如发送“YL 苏新宁-10-3”即查询苏新宁第10篇被引之第3篇来源信息,过程如上,不同的是调用了作者被引来源函数,结果以标准的著录格式给出,其中书籍、期刊论文、报纸、会议录等不同文献种类有不同的格式。如期刊论文形为:

邓三鸿.电子政务技术框架.现代图书情报技术.2004,(01):72-75.联系电话:*****。

以上列举的是简单的查询,有些功能的查询式构造虽然简单,但结果数据量有时很大,如全文、图片,全部通过短信分条发送可能并非最佳选择,可以让用户发送E-mail地址至系统,系统自动将查询结果发送到用户邮箱中。

3.4 其它服务及发展方向

除了提供信息查询外,由于对每一次的数据库访问都有相应的日志记录,其中包括用户的手机号码、查询串、查询时间等一系列完整的信息。我们不仅可以随时查看某时段的日志,而且系统经过一段时期的使用,必然有大量的记录积累,于是我们可以通过数据挖掘的若干方法,从记录集中寻找针对某一用户查询的一般性规律,进而不难对不同用户提供个性化的服务,还可以进一步实施信息推介,将最新用户感兴趣的更新或动态信息主动推送给用户,变被动接受请求为主动提供信息,提高服务的质量和层次。例如发现某一用户查询集之中一半甚至更多都是关于某一著者或主题的内容,我们就可以认为用户对该主题或著者兴趣度较大,有这方面的最新信息,我们就可以及时主动通知该用户;还有,在用户查找的关键词中,通过相似度向量计算发现这些词汇基本都

集中在某一个或几个学科主题上,同样我们可以将该用户列为这一个或几个学科主题的主要服务对象之一。对于相同或类似的服务对象,为减少工作量,可以建立群组,实现群发,甚至可以基于本系统开辟他们彼此之间的讨论平台,充分实现信息共享。这类深层次的服务有待于以后逐步去实现,它也是短信平台服务的发展方向。

4 结 语

ISAPI以其独特的优点而广泛应用于Web服务中,籍于此的CSSCI短信平台随着功能的逐步完善和服务方式的多样化发展,不仅是对网络检索的成功扩充,对其他短信平台建设也起到良好的借鉴意义。同时,利用短信平台提供查询和更深层次的服务,应努力做到以用户需要和方便为发展动力,从用户角度考虑系统升级和功能改进,充分体现“以用户为本”的思想。

参考文献:

- 1 苏新宁.中国社会科学引文索引设计.情报学报,2000,19(4):290-295
- 2 马费成.CSSCI与社会科学评价.南京大学学报(哲学·人文科学·社会科学),2000,37(4):155-160
- 3 Tony Beveridge, Paul McGlashan. ISAPI/NSAPI WEB 高级编程.北京:水利水电出版社,1999
- 4 范刚峰,江润莲.用ISAPI Filter建立防火墙.计算机工程,2002,28(8):193-194,203
- 5 党京,孙若雯,于蓉蓉等.中国移动通信互联网短信网关接口协议. Oct. 15, 2004. <http://down.upsdn.net/telecom/ChinaMobile-SMS-GateWay-Protocol.doc>. (Accessed Jun. 10, 2005)
- 6 汤敏华,曹健.基于短信的信息服务平台研究.计算机工程,2004,30:238-240.

(作者E-mail:zcmatrix@163.com)

(上接第78页)

- 3 NT Ozok, Gavriel Savendy. Twenty guideline for the design of Web-based interface with consistent language, 2004
- 4 Pam L. Kumar, Michael Alan Smith, Snehamay Bannerjee. User interface features influencing overall ease of use and personalization, 8 June 2003
- 5 Hock-Hai Teo, Lih-Bin Oh, Chunhui Liu, Kwok-Kee Wei. An empirical study of the effects of interactivity on Web user attitude, 23 December 2002
- 6 G. Bhatt. Bringing virtual reality for commercial Web sites, 12 July 2003
- 7 Hazel Hall, Dianne Graham. Creation and recreation: motivating collaboration to generate knowledge capital in online communities, June

2004

- 8 Herbjorn Nysveen, Per E. Pedersen. An exploratory study of customers' perception of company Web sites offering various interactive applications: moderating effects of customers' Internet experience, 4 December 2002
- 9 Su-e Park, Dongsung Choi, Jinwoo Kim. Critical factors for the aesthetic fidelity of Web pages empirical studies with professional Web designers and users, 14 July 2003
- 10 Wilson Lee, Izak Benbasat. Designing an electronic commerce interface: attention and product memory as elicited by Web design, 18 April 2003

(作者E-mail:fanghainlin@sjtu.edu.cn)