

# 基于 Ajax 长轮询技术的在线答疑系统设计\*

## Online Q&A System Design Method based on Ajax Long-polling

唐昭琳<sup>1</sup>, 顾林<sup>2</sup>

TANG Zhao-lin<sup>1</sup>, GU Lin<sup>2</sup>

(1. 广西师范学院, 广西南宁 530023; 2. 南宁市平方软件新技术有限责任公司, 广西南宁 530004)

(1. Guangxi Teachers Education University, Nanning, Guangxi, 530023, China; 2. Nanning Pingsoft New Technology Co. Limited, Nanning, Guangxi, 530004, China)

**摘要:**设计一种基于 Ajax 长轮询技术来实现服务器消息推送的在线答疑系统,并对系统的并发性能进行优化设计。与传统的客户端轮询服务器方式相比较,基于 Ajax 长轮询技术在线答疑系统的负载能力要明显大些。

**关键词:**在线聊天 Ajax 长轮询 消息推送

中图分类号: TP391 文献标识码: A 文章编号: 1002-7378(2013)01-0040-03

**Abstract:** A new approach is designed for Online Q & A System based on Ajax long-polling to achieve messages pushing from server to clients. The concurrent performance of the system has been optimized and experimental results show that the proposed method has better performance than traditional client's poll method.

**Key words:** online chat, Ajax long-polling, message pushing

基于 Web 的在线答疑系统是网络教学系统的重要组成部分。在线答疑系统可为师生之间提供一个互动式的在线交流平台。学生可以向线上的辅导老师提问,辅导老师可以通过网络及时解答学生疑问。在线答疑系统中,两个客户端之间的消息收发通过服务器进行,一旦有一方客户端发出对话消息,即要求服务器能正确、及时地将此消息推送至接收消息的另一方客户端,即服务器能实时地将更新的信息推送到接收消息的客户端。传统的 Web 应用系统的工作方式是浏览器发出请求、服务器响应的 Web “请求—响应”模式,浏览器的主要工作是发送请求、解析服务器返回的信息以不同的风格显示,这样每个客户端为了得到最新的消息,都要主动、定时地向服务器“轮询”,发出访问请求,如发现有新消息即更新页面。当来自客户端的并发访问数量变大

时,这种方式无疑会给服务器带来很大压力,不适合高并发用户量的应用。本文提出一种基于 Ajax 长轮询技术的在线答疑系统,主要实现在线讨论和答疑功能,在系统中学生或老师可以创建临时讨论组、共享文件资料、查看以往讨论内容,还可以管理自己的通讯录等。支持一对一和一对多(群)两种类型的在线讨论。交流内容可以是图片、文字等。客户端不需要下载和安装任何插件。

## 1 系统设计

### 1.1 软件架构设计

系统软件采用 B/S 架构、MVC 模式,分为表示层(View Layer)、业务层(Business Layer)和数据层(Data Layer)。采用 SSH(Struts + Spring + Hibernate)的组合框架模式开发,并利用 DWR、ExtJS、Json 等技术实现系统的开发。系统软件架构如图 1 所示。

收稿日期:2012-06-12

修回日期:2013-01-06

作者简介:唐昭琳(1965-),男,高级工程师,主要从事多媒体和专家系统设计开发工作。

\*广西科技计划创新能力建设项目(桂科能 05112001-7C)资助。

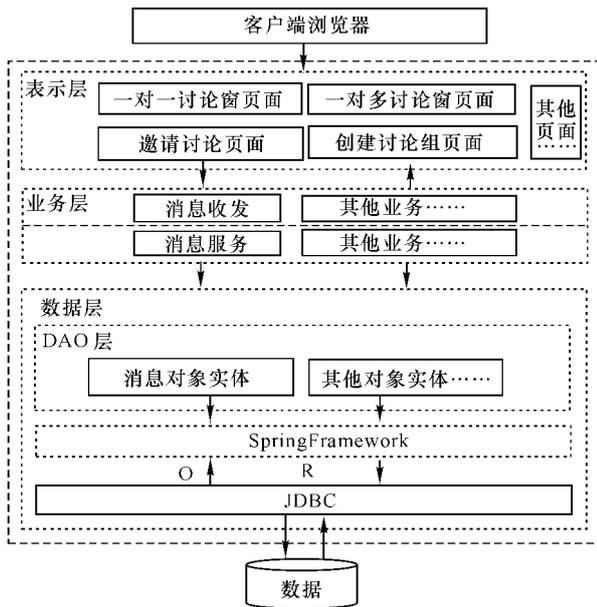


图 1 系统软件架构

表示层由 js 组成,负责客户端的展示和交互,包括窗口的创建、显示等。主要采用 ExtJS 框架实现界面。业务层采用 DWR 框架;DWR 是一个用于改善 web 页面与 Java 类交互的远程服务器端 Ajax 开源框架;通过 DWR 可以在客户端利用 JavaScript 直接调用服务端的 Java 方法并返回值给 JavaScript,如同本地客户端直接调用。数据层采用 Hibernate Framework 进行 OR 映射,隔离了数据访问细节,方便移植到不同的数据库系统下。

### 1.2 消息推送的设计

系统采用基于 HTTP 长连接的 Ajax (Asynchronous JavaScript and XML) 长轮询来实现服务器消息推送。Ajax 长轮询是指客户端在 XMLHttpRequest 的 readystate 为 4 (即数据传输结束) 时调用回调函数,进行信息处理。当 readystate 为 4 时,数据传输结束,连接已经关闭。相对于客户端主动“轮询”,这种方式也可以称为“拉”(pull)。该方式具有以下一些优点:请求异步发出;无须安装插件;IE、Mozilla FireFox 都支持 Ajax。使用 Ajax 长轮询实现服务器消息推送与传统的 Ajax 应用不同之处在于:(1)服务器端会阻塞请求直到有数据传递或超时才返回。(2)客户端 JavaScript 响应处理函数会在处理完服务器返回的信息后,再次发出请求,重新建立连接。(3)当客户端处理接收的数据、重新建立连接时,服务器端可能有新的数据到达;这些信息会被服务器端保存直到客户端重新建立连接,客户端会一次把当前服务器端所有的信息取回。

系统采用 DWR(Direct Web Remoting) 框架实

现 Ajax 长轮询。DWR 是一个 Web 远程调用框架。DWR 包含以下两个主要部分:(1)一个运行在服务器端的 Java Servlet,它处理请求并且向浏览器发回响应;(2)运行在浏览器端的 JavaScript,它发送请求而且还能动态更新网页。利用 DWR 推模式向所有客户端推送数据,只要客户端的访问页面确定就可以了。但是,若需要对访问同一页面的两个客户端,特别是根据用户身份推送不同数据时,就必须加以区分客户端。我们设计系统时对每个客户端用 ScriptSession 给予唯一标识的方式来区分,即将所有的 ScriptSession 和对应的 SessionID 用 map 保存于服务器端,服务器端根据 SessionID 取得 ScriptSession 向该客户端推送数据时,根据这个属性加以过滤。这样在一对一或一对多的在线对话中,服务器就能根据客户端的 ScriptSession 选择消息接收方进行推送。

### 1.3 优化性能设计

Ajax 长轮询是基于 HTTP 长连接的。与一般每次发起 HTTP 请求或响应都要建立一个 TCP 连接不同,HTTP 长连接利用同一个 TCP 连接处理多个 HTTP 请求和响应,也叫 HTTP keep-alive,或者 HTTP 连接重用。使用 HTTP 长连接可以更少地建立和关闭 TCP 连接,从而减少网络流量,提高 HTTP 请求/响应的性能。

对于一个实际的应用而言,系统的稳定性和性能是非常重要的。将 HTTP 长连接应用于实际,一般 Web 服务器会为每个连接创建一个线程,在大量并发用户访问下,服务器端需要维护大量并发的长连接。在这种应用背景下,除了服务器端需要考虑采用负载均衡和集群技术,我们在服务器端为长连接作了改进:(1)不同类型的信息使用不同类型的连接。使用长连接时,存在一个很常见的场景:客户端网页需要关闭,而服务器端还处在读取数据的堵塞状态,客户端需要及时通知服务器端关闭数据连接。服务器在收到关闭请求后首先要从读取数据的阻塞状态唤醒,然后释放为这个客户端分配的资源,再关闭连接。所以在设计上,对客户端的控制请求和数据请求使用不同的 HTTP 连接,以控制请求不会被阻塞:客户端异步地发出一个 XMLHttpRequest 请求,通知服务器端关闭数据连接。(2)在客户端和服务器之间保持“心跳”信息。在浏览器与服务器之间维持一个长连接会为通信带来一些不确定性,因为数据传输是随机的,客户端不知道何时服务器才有数据传送。服务器端需要确保当客户端不再工作

时,释放为这个客户端分配的资源,防止内存泄漏。因此需要一种机制使双方知道大家都在正常运行。我们在实现上采用两种方法:①服务器端返回数据、关闭连接后,经过某个时限没有收到客户端的再次请求,则认为客户端不能正常工作,服务器端释放为这个客户端分配、维护的资源。②当服务器处理信息出现异常情况,需发送错误信息通知客户端,同时释放资源、关闭连接。

## 2 系统测试

系统性能测试的服务器硬件环境为:Web 服务器和数据库服务器 1 台,CPU 为 AMD Athlon(tm) 64 X2 Dual 2.4GHz,2G 内存。测试的服务器软件环境为:操作系统 Window Server 2003,数据库系统 MySQL 5.1 版本,Web 应用服务器采用 Tomcat 6.0.20。测试要点:(1)支持的最大在线人数;(2)并发聊天消息的响应速度。在长轮询方式下,测试时分别模拟 200 个用户、500 个用户同时登录的聊天室,每个用户以 1 条消息/秒的速度发送消息。在客户端轮询方式下,测试时分别模拟 50 个用户、100 个用户同时登录的聊天室,每个用户以 1 条消息/秒的速度发送消息。对比测试后结果如表 1 所示。

从表 1 的测试结果看,第 1 种方式 200 个用户并发时系统延迟速度属于可接受范围,500 个用户

并发时系统延迟较多。第 2 种方式下 50 个用户并发时系统延迟速度属于可接受范围,100 个用户并发时系统延迟较多。可见,第一种方式优于第二种方式。

表 1 测试结果

消息推送方式	用户数	线程数	发送平均延迟(ms)	接收平均延迟(ms)
基于长轮询的服务器推送	200	200	1095	662
	500	500	3050	3078
客户端轮询	50	50	1528	781
	100	100	4014	3896

## 3 结束语

本文设计的在线答疑系统,使用基于 HTTP 长连接的 Ajax 长轮询实现消息推送,采用 DWR 框架实现,测试结果证明,与传统的客户端轮询服务器方式相比较,系统的负载能力要明显大些。但是在刷新页面时会产生一个新的 ScriptSession 实例,造成服务器端与客户端的原有连接丢失,从而无法将消息推送到客户端,因此,系统的稳定性有待进一步提高。

(责任编辑:邓大玉)

(上接第 39 页)

- [5] 于春雪. 基于 STM32F107 的高速以太网接口设计与应用[J]. 网络与多媒体, 2011(9): 63-67.
- [6] 宋鑫, 郭勇, 谢兴红. RMI 模式以太网 PHY 芯片 DP83848C 的应用[J]. 单片机与嵌入式系统应用, 2010(8): 50-51.
- [7] Adam Dunkels. Design and implementation of the LwIP TCP/IP stack[M]. Swedish Institute of Computer Science, 2001.
- [8] 孙乐鸣, 江来, 代鑫. 嵌入式 TCP/IP 协议栈 LWIP 的内部结构探索与研究[J]. 电子元器件应用, 2008, 10

(3): 79-82.

- [9] 焦海波, 刘健康. 嵌入式网络系统设计[M]. 北京: 北京航空航天大学出版社, 2008: 248-285.
- [10] 孔栋, 郑建宏. 嵌入式 TCP/IP 协议栈 LWIP 在 ARM 平台上的移植和应用[J]. 通信技术, 2008, 41(6): 38-40.
- [11] 万文, 熊震宇. 基于 LabVIEW 的焊接过程分析系统软件设计[J]. 电焊机, 2010, 40(3): 43-45.

(责任编辑:邓大玉)