

Mobile Educational Administration Query System Based on HTTP Parsing and Simulation Techniques

Haijiao Xu, Zhanhong Zhang, Yilong Zhou

Department of Computer Science, Guangdong University of Education, Guangzhou Guangdong
Email: guesskkk99@163.com, 206080110@qq.com, 1397065979@qq.com

Received: July 31st, 2019; accepted: August 15th, 2019; published: August 22nd, 2019

Abstract

In the mobile Internet era, in order to tackle that the issue of conventional educational administration query system was not compatible with mobile access based on mobile phones and tablet computers, this paper designed a mobile educational administration query system based on the existing educational administration system. It is an extension of conventional campus educational administration system to mobile intelligence system. We first investigate the HTTP parsing and simulation techniques, including HTTP request parsing, HTTP response parsing and HTML message body parsing. Then, taking the HTTP student score query as an example, we introduce the Java development process based on these HTTP parsing and simulation techniques. The key design ideas and procedures are analyzed and provided, which can provide the guidance for the actual development of a multifunction mobile educational administration query system.

Keywords

HTTP Parsing, Simulation Login, Mobile Educational Administration Query System

基于HTTP分析模拟的移动教务查询系统研究

徐海蛟, 张展鸿, 周奕龙

广东第二师范学院, 计算机科学系, 广东 广州
Email: guesskkk99@163.com, 206080110@qq.com, 1397065979@qq.com

收稿日期: 2019年7月31日; 录用日期: 2019年8月15日; 发布日期: 2019年8月22日

摘要

在移动互联网时代下, 针对传统的高校教务查询系统不兼容手机和平板等设备的移动访问问题, 在不变

更已有PC版教务系统的基础上分析并设计了一种简单便捷的移动教务查询系统,这也是移动时代下传统校园教务系统向移动智能终端的延伸。首先,研究了HTTP请求分析、HTTP响应分析、HTML消息体解析等HTTP分析模拟技术及方法,然后以教务系统的HTTP成绩查询为例,阐述基于HTTP分析模拟技术的Java开发过程,重点给出HTTP学生成绩查询实现过程中的关键设计思想及步骤,为最终开发实现多功能教务信息移动查询系统提供了重要指导。

关键词

HTTP分析, 模拟登录, 移动教务查询系统

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着我国高等教育的快速发展,教务管理工作成为高校日常校务的重要环节,高校教务查询系统的建设也成为完善高校教务信息管理的必经之路[1]。传统的高校教务信息查询系统开发时期较早,存在着诸多不便:登录使用流程复杂;仅支持PC端浏览器访问而不兼容手机和平板等移动端访问,移动适配性和兼容性较差;访问地域受限制,只支持校园内网访问等,这影响了在校的师生对教务查询系统的使用体验。高校师生经常使用教务信息系统去查询成绩、课表等信息,为了提升高校师生使用教务查询系统的体验,支持师生们随时随地的移动查询需求,设计一种便捷的移动教务查询系统势在必行。该移动系统满足几个基本需求:支持教务信息的移动查询访问、保持已有教务系统无变更情况下的无缝接入,它是移动互联网时代下传统B/S架构校园教务系统向移动智能终端的延伸。

HTTP是Hyper Text Transfer Protocol(超文本传输协议)的缩写,它的发展是万维网协会和Internet工作小组IETF(Internet Engineering Task Force)合作的结果,最终发布了一系列的HTTP RFC标准版本,其中最著名的就是RFC 2616标准版。RFC 2616定义了普遍使用的一个版本——HTTP 1.1。HTTP是一个属于应用层的面向对象的TCP/IP协议,由于其简捷、快速的方式,适用于分布式超媒体信息系统。HTTP协议的主要特点可概括为:支持B/S模式;服务简单高效;数据对象类型灵活;无连接通信方式;无状态工作协议。HTTP是一个基于请求与响应模式的、无状态的应用层协议,常基于TCP的连接方式,目前大多数的Web开发是构建在HTTP协议之上的Web应用。高校使用的教务信息系统主要是正方教务等HTTP Web应用[2],HTTP分析模拟技术为实现这种Web查询系统提供了技术基础。通过对该HTTP Web教务系统的登录与查询等流程进行HTTP分析,解析提取HTML(Hypertext Markup Language)页面中用户需要的关键数据,经数据的清洗和整理,最后返回查询结果给移动端用户。基于这种HTTP分析模拟技术设计的教务查询代理系统,无需变更现有的教务系统架构设计,也不影响教务管理的正常运行,通过Android API[3]或微信小程序[4]等移动开发技术提供对移动端的支持,拓展了教务查询系统的移动版,解决了用户使用流程繁琐、不匹配移动端等痛点需求。

2. HTTP分析模拟方法

2.1. HTTP请求分析

HTTP协议是由Web客户端发起请求,服务器回送响应。一次HTTP交互操作称为一个事务,其工

作过程可分为四个步骤。以 Web 教务查询为例, 1) 首先客户机与 Web 服务器需要建立连接。当输入教务系统的 URL (Uniform Resource Locator)网址或单击其超链接, 则 HTTP 交互工作开始; 2) 客户机发送一个请求给教务系统服务器, 请求方式的格式为: URL 网址、协议版本号、MIME 信息(包括请求修饰符、客户机信息和其他可能的内容); 3) 教务系统服务器给予相应的响应信息, 其格式为一个状态行, 包括信息的协议版本号、一个成功或错误的代码, 后边是 MIME 信息包括服务器信息、实体信息和查询内容; 4) 客户端接收 Web 服务器所返回的结果数据并通过浏览器界面显示, 最后客户机与教务系统服务器断开连接。

HTTP 请求由三部分组成, 分别是: 请求起始行、请求消息头、请求消息体[5]。请求起始行以一个方法符号开头, 以空格分开, 后面跟着请求的 URI 和协议的版本, 格式如下: Method Request-URI HTTP-Version CRLF, 其中 Method 表示 HTTP 请求方法如 GET 或 POST 等; Request-URI 是一个 URL 访问网址; HTTP-Version 表示请求的 HTTP 协议版本如 HTTP/1.1; CRLF 表示回车和换行。请求消息头允许浏览器客户端向服务器端传递请求的附加信息以及客户端自身的信息, 常见的请求头域有 Accept (浏览器可接受的 MIME 类型)、Referer (用户从该 URL 页面出发访问当前请求的页面)、User-Agent (浏览器类型, 如果返回的内容与浏览器类型有关则该值很有用)等。HTTP 请求消息体含有其他附加信息, 移动端常见的请求体是任意类型, Web 服务器不解析消息体, 消息体的处理需由上层应用自行解析。

2.2. HTTP 响应分析

在接收和解释请求消息后, 服务器返回一个 HTTP 响应消息。HTTP 响应也是由三个部分组成, 分别是: 响应状态行、响应消息头、响应消息体。响应状态行格式如下: HTTP-Version Status-Code Reason-Phrase CRLF, 其中, HTTP-Version 表示服务器 HTTP 协议的版本; Status-Code 表示服务器发回的响应状态代码; Reason-Phrase 表示状态代码的文本描述。响应状态行包括了 HTTP 协议版本、HTTP 状态码和状态等重要信息, 其中 HTTP 状态码是判断请求结果的重要标准。响应消息头允许服务器传递不能放在响应状态行中的附加响应信息, 以及关于服务器的信息和对 Request-URI 所标识的资源进行下一步访问的信息。它包含各种属性, 其中 Cookie 信息是较为重要的一个标记用户的属性, Cookie 能让网站“记住”访问者的身份, 如若有了 Cookie, 用户就不需要每次都输入密码进入系统, 是 Web 应用保存用户身份的常见工具。响应消息体是服务器返回的 HTML 正文信息, 通过解析响应消息体中的 HTML 就能获取需要的数据信息。

用户在使用教务系统应用时, 通过多次这样的 HTTP 交互过程, 完成教务 HTML 应用数据的传递和查询。因此, 可以通过对交互过程中的浏览器向教务系统发出的 HTTP 请求以及教务系统返回的响应消息体进行 HTTP 分析, 探究教务系统 Web 应用交互的规律。利用这种 HTTP 分析模拟技术, 模拟用户发送 HTTP 请求到教务 Web 系统, 并对 Web 教务系统 HTML 格式的响应消息体进行解析、清洗、定位, 获取用户需要的目标数据, 最终呈现于移动端用户界面, 完成基于分析模拟的 HTTP 交互流程。

2.3. HTML 消息体解析

Web 教务系统服务器返回的 HTTP 响应消息体是一个 HTML 半结构化网页数据。HTML 是用来描述多媒体网页的一种超文本标记语言, 包含很多由尖括号包围的 HTML 标记, 其含有所需的教务应用数据。HTML 语法是由 W3C 组织创建的规范中进行定义的, 其语法格式是由 DTD (Document Type Definition) 定义的, 该格式定义了允许的元素、属性和层次结构。因为 HTML 响应数据的包容性语法特性, 以及在解析过程中存在动态脚本如 document.write()会改变 HTML 内容, 导致无法使用自上而下或是自下而上的

解析。HTML 解析过程首先是词法分析即 HTML 标记化, 整体算法的核心是解析状态机的改变, 即解析过程中由状态机标识当前的状态到了哪一个解析阶段。其次, 构建文档对象模型(Document Object Model, 简称 DOM)树[6], 将对应的 HTML 标记添加到 DOM 树上, 此过程中也有一个状态机去维护对应的阶段。DOM 树是 HTML 文档的映射关系和 HTML 元素对外的接口, 每一个节点是由 DOM 元素和节点属性组成。如果遇见内部<script>标签, 就会中断 HTML 解析, 开始执行脚本, 直到脚本执行完毕, 而遇到外部的脚本, 解析中断直到请求脚本回来, script 解析模式在 HTML 4/5 规范中有详细描述。对于风格样式解析, 由于解析样式不会影响 DOM 树构建, 也不包含关键的教务数据, 所以文中不予考虑。我们的移动教务查询系统采用, Jsoup [7]是一款基于 Java 的 HTML 解析器, 可解析某个 URL 网址和 HTML 文本内容。通过 DOM 树以及近似 jQuery [7]的简易操作方法取出教务数据, 其操作步骤可归为三步: 1) 基于 HTTP GET 请求获取对应 HTML 页面数据; 2) 利用 Jsoup 把 HTML 页面数据解析为 Document 对象; 3) 利用 Document 对象的 getElementById()、getElementsByTag()、select()等定位器方法依据教务数据特征定位获取目标教务数据。

3. 移动教务查询系统开发——以 HTTP 成绩查询为例

以我校正方教务系统的 HTTP 成绩查询为例, 讲述基于 HTTP 分析模拟技术的 Java 开发过程, 重点给出 HTTP 学生成绩查询实现过程中的关键设计思想。移动端网络框架 Java OKhttp [8]是一个适于处理 HTTP 交互操作的轻量级网络请求框架, 它为实现移动教务查询系统提供了技术基础。不失一般性, 其他类型的教务系统如青果教务系统等[9]的设计及实现原理基本相同。

3.1. HTTP 模拟登录开发

移动教务查询系统首先需要验证学生身份, 包含一个 HTTP 模拟登录模块。通过对用户登录的 HTTP 交互报文分析, 发现浏览器向中央认证服务器以 POST 的方式发送了一个 HTTP 请求。教务查询系统 HTTP 请求消息体中的 HTML 表单传递了 username、password 和 service 三个关键参数, 其中, username 代表用户的账号名, 即在校学生的教务系统账号; password 代表该账号的密码; service 是一个固定的网址, 表示登录到教务系统的管理页面。假若用户名及密码验证成功, 认证服务器会返回一个含有 JavaScript 脚本的页面, 它重定向到带有 ticket 属性的网址。由于 HTTP 模拟登录技术采用的移动端 Java OKhttp 网络框架不会自动执行 JavaScript 脚本, 因此需要对 JavaScript 脚本进行内容分析以确定重定向网址, 之后模拟执行 URL 跳转。首先, 通过 Java Jsoup 组件获取此页面中 JavaScript 的网络地址 URL, 然后通过 HTML GET 方法访问 URL 网址, 从而实现模拟重定向页面跳转。假若用户名及密码验证失败, 则返回登录失败页面, 此时可在用户界面上提示用户名密码错。

模拟重定向访问 JavaScript 所含 URL 网址后, 教务服务器会返回一个状态码为 302 (重定向)的 HTTP 响应报文, 二次重定向地址正是该报文的 HTML a 标签中的 URL 地址。由于该二次重定向行为不是 JavaScript 脚本执行的, 所以需要根据 Java OKhttp 网络框架的重定向自动与否属性设置来判断是否需要通过 HTML GET 访问该 HTTP 响应报文中 a 标签所含有的 URL 网址。二次重定向后, 返回一个含有 JavaScript 脚本的 HTML 页面, 分析该 JavaScript 脚本的含义, 发现其将打开一个路径为/xs_main.aspx?xh=(/学号)的 HTML 页面。因此, 在 OKhttp 网络框架中使用 GET 方法打开该 URL 网址, 第三次 URL 重定向而进入教务系统个人主页, 完成 HTTP 模拟登录。假若返回的 HTML 页面 head 标题为“正方教务管理系统”, 且可以在 id 为 xhxm 的 span 标签中发现学生的姓名, 则表明移动端 HTTP 模拟登录成功。登录进入个人主页后, 可缓存用户个人信息到个人 Session 对象中, 以便跨功能模块使用, 整个流程如图 1 所示。

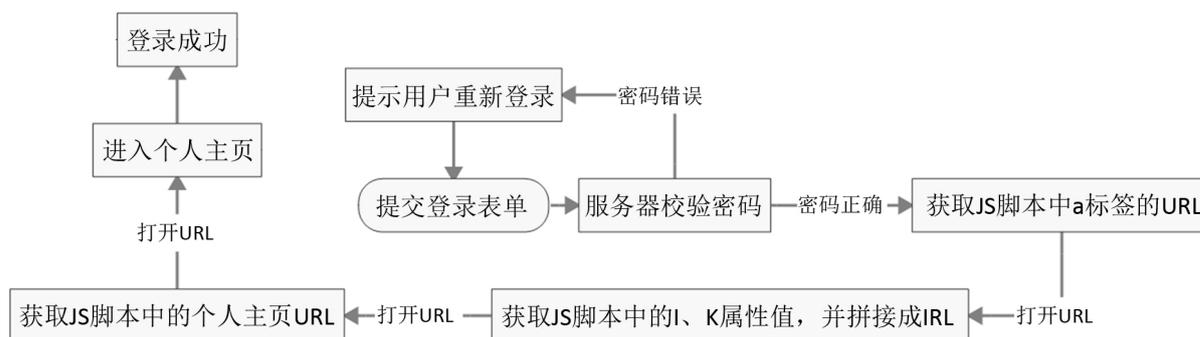


Figure 1. HTTP simulation login flowchart

图 1. HTTP 模拟登录流程图

3.2. HTTP 成绩查询开发

查询学期课程成绩是学生最常用的教务功能之一。HTTP 成绩查询的前置条件是已登录。当移动端完成 HTTP 模拟登录后，在 Session 对象中保存着用户名、密码等个人信息，可直接获取 Session 对象中的个人信息进行查询。假若用户未登录，Java 拦截器会拦截查询请求，并强制跳转到登录界面，要求用户执行 HTTP 登录。通过分析成绩查询的 HTTP 交互信息，发现浏览器向服务器以 POST 方式发出了一个 HTTP 请求，传递了 ddIXN、ddIXQ、Button1、_VIEWSTATE 四个 HTML 关键表单参数。此外，还以 GET 查询串方式向服务器提交了 xh、xm 和 gnmkdm 三个 HTML 关键参数。

HTML 表单参数 ddIXN 和 ddIXQ 分别表示成绩查询的学年和学期。HTML 表单参数 Button1 属于固定值，用于服务器标识查询的内容。_VIEWSTATE 表单参数是 ASP.NET 的视图状态值，用于保存会话的 Base64 编码值，在 HTML 报文中是一个隐藏域表单数据，表示 HTTP 交互过程中需要向 Web 服务器中传递的 _VIEWSTATE 参数。GET 查询串参数 xh 和 xm 分别表示学号和姓名，对于中文姓名，通过网络传递 UTF-8 类型的中文字符，会导致乱码。因此，HTML 使用 URLEncoder 将 xm 参数进行 URL 编码，转换为 application/x-www-form-urlencoded MIME 字符串来传递。GET 查询串参数 gnmkdm 作用同 Button1。

在教务系统返回的 HTTP 响应报文中，查找 class 属性为 datelist 的 table 标签，含有对应学期、学年的成绩信息。每个 tr 标签包含一个课程成绩信息，每个 td 标签包含一个基本单元课程成绩信息。利用 Java Jsoup 可解析和定位目标成绩数据。首先使用 Jsoup select 选择器查找 _VIEWSTATE 参数值，然后向教务系统以 POST 方式提交查询表单，发出 HTTP 成绩查询请求。最后，系统返回成绩表响应报文，通过 Jsoup 定位 HTML 成绩表格，获取表格所有 tr 元素，循环遍历每个 tr 元素，可得课程成绩数据表。

RESTful API 是一个 API 设计规范，适于 HTTP 成绩查询这类轻量级的前后端数据接口设计，Java 技术实现 RESTful 接口有三种方法：HttpClient 实现、URLConnection 实现、SpringRestTemplate 实现 [10]。文中采用简洁的 HttpClient 实现 RESTful API 以完成移动客户端与 Java 控制器端的代理数据交换。为提高查询性能，考虑将查询获取的成绩数据缓存到高速 NoSQL [11] 数据库中，当 MongoDB 数据库有该学生成绩缓存时则直接高速取出使用，当成绩数据不在 MongoDB 数据库中时则通过 Web 教务服务器取出学生成绩数据并缓存到 MongoDB 数据库。选择 MongoDB 这种基于分布式存储的高速 NoSQL 数据库来储存成绩数据，主要考虑是：1) 学生批量查阅成绩往往集中在学期末放假前，易产生成绩查询的高并发瞬时峰值，MongoDB 数据库可支持大规模数据存储并较好支持高并发低延时的 Web 查询，提高用户响应度和用户查询体验；2) 成绩记录一旦产生，通常没有二次修改，适于成绩查询缓存；3) 高校教务查询系统应用的复杂度不高，对数据完整性约束要求不强，适用于 MongoDB 这种新型的、快速的 NoSQL 数据库系统。

3.3. HTTP 成绩查询测试

HTTP 成绩查询的前置条件是用户已登录,当 Java 拦截控制器检测发现 Session 对象未含有用户登录信息,则拦截查询请求之后强行跳转至登录页面,要求移动客户执行 HTTP 登录。当移动端已完成 HTTP 模拟登录,Java 控制器可直接提取保存于 Session 对象中的用户信息进行成绩查询。由于成绩查询 RESTful API 返回的学生成绩数据格式,不是纯文本而是一个 Json 通用数据交换格式[12],因此,Web 服务器响应的 HTTP 消息头的 Content-Type 属性设置为 application/json。使用 PostMan HTTP 测试器工具测试成绩查询 API 接口,返回的一个 JSON 格式的成绩数据示例如图 2 所示,成绩列表 GradeList 属性中存放着多个 Grade 信息元素:学年、学期、课程名、学分、课程类型、成绩等。

通过 Android SDK API 或微信小程序 API 等移动开发技术提供对移动端开发支持,将返回的 Json 格式成绩列表 GradeList 以简洁便利的方式展示在移动终端如手机等,查询效果如图 2 所示。



我的成绩单			
大一	大二	大三	大四
第一学期 平均学分绩点: 3.22			
课程	学分	成绩	
Java技术	4.0	91	
操作系统	4.0	81	
代码重构与设计模式	3.0	83	
分布式计算技术	4.0	82	
概率论与数理统计	2.0	78	
计算机网络	4.0	84	
数据库系统应用	4.0	75	
数据库系统原理课程设计	1.0	70	
网页设计技术	3.0	85	
形势与政策 V	0.0	86	
学科竞赛 III	1.0	84	
职业认证 I	1.0	83	

Figure 2. An example of student score query

图 2. 学生成绩查询示例

4. 结束语

为支持高校教务信息移动查询需求,设计了一种基于 HTTP 分析模拟技术的移动教务查询系统。文中以 HTTP 成绩查询为实例,讲解了 HTTP 分析模拟程序的关键思路与方法。类似地,还可开发 HTTP 课表查询、HTTP 校园卡消费查询以及 HTTP 图书借阅查询等,最终实现多功能教务信息移动查询系统。

基金项目

2019 年广东第二师范学院校级教学质量与教学改革工程项目(编号: 2019jxgg18); 广东第二师范学院软件工程重点学科建设项目(编号: 9030-1700207); 广东省自然科学基金项目(编号: 2018A0303130169); 广东省科技计划项目(编号: 粤财农[2017]94 号, 2016A010106007); 广东省应用型科技研发专项资金资

助项目(编号: 2016B090927010); 2019 年广东第二师范学院大学生创新创业训练计划项目(编号: 201914278146)。

参考文献

- [1] 张明. 基于实践教学视角的高校教学教务管理信息化能力提升研究[J]. 中国管理信息化, 2018, 21(14): 222-223.
- [2] 马浩东. 高校正方教务系统成绩一键导入与下载的设计与实现[J]. 赤峰学院学报(自然科学版), 2018, 34(8): 63-65.
- [3] 毛东航, 马思明, 毛姝然, 张金霞. 基于 android 的教务管理系统[J]. 电子世界, 2017(19): 73+75.
- [4] 吴政修. 微信小程序在教务服务中的应用[J]. 信息系统工程, 2018(5): 172-173.
- [5] 赵琦, 蒋朝惠, 周雪梅, 宋紫华. 一种基于 HTTP 协议的隐蔽隧道及其检测方法[J]. 计算机与现代化, 2019(6): 16-23+29.
- [6] 赵建视, 梁俊卿, 吕笑琳, 黄学波, 冷悦, 张在军. 基于 DOM 树统计关键词路径的 WEB 信息提取研究[J]. 计算机科学与应用, 2019, 9(2): 181-187.
- [7] 毛凯. 基于 Jsoup 的通用网页采集系统的设计与实现[D]: [硕士学位论文]. 成都: 电子科技大学, 2015.
- [8] 李群. 基于 OkHttp 的文件传输设计与实现[J]. 电子技术与软件工程, 2018(13): 180-181.
- [9] 闫春梅. 现代教育管理与现代信息技术视角下的“教务管理系统”[D]: [硕士学位论文]. 呼和浩特: 内蒙古师范大学, 2009.
- [10] 刘香利. 基于 RESTful/HTTP 的网络管理接口定义方法和通知机制设计[D]: [硕士学位论文]. 北京: 北京邮电大学, 2019.
- [11] 祝幸辉. 云物联平台的数据汇聚与存储方法研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2017.
- [12] 裴妍. 基于 Android 的移动教务平台的设计与实现[D]: [硕士学位论文]. 长春: 吉林大学, 2016.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org