

大数据背景下BBS的设计

韩中¹, 苟和平¹, 梁栋¹, 韩玟璐², 唐国纯¹, 赵安学¹

¹琼台师范学院, 信息科学技术学院, 海南 海口

²海南大学, 应用科技学院, 海南 海口

收稿日期: 2022年3月8日; 录用日期: 2022年4月8日; 发布日期: 2022年4月15日

摘要

针对大数据的数据源—BBS (Bulletin Board System)的构建问题, 提出了大数据背景下BBS的设计研究。研究突破传统的BBS系统设计, 采用了全新的多层Web架构, 支持大数据云服务的Nginx控制方式。使用Python工具来实现接口数据的快速表达, 设计多主题的功能模块满足最新用户的需求。基于上述思想, 详细地介绍了用户的需求提炼、功能模块的划分、数据模型设计、技术构架设计, 以及重要的接口设计等方面的内容。通过BBS论坛的全新设计, 使人们能够在更多的平台上面进行交流, 用户间的沟通将会变得更加快捷、便利, 为大数据分析提供更加丰富的数据资源。

关键词

BBS, 大数据, Web架构, 数据源

Design of BBS under the Background of Big Data

Zhong Han¹, Heping Gou¹, Dong Liang¹, Wenjun Han², Guochun Tang¹, Anxue Zhao¹

¹School of Information Science and Technology, Qiongtai Normal University, Haikou Hainan

²School of Applied Science and Technology, Hainan University, Haikou Hainan

Received: Mar. 8th, 2022; accepted: Apr. 8th, 2022; published: Apr. 15th, 2022

Abstract

Aiming at the construction of BBS (Bulletin Board System), the data source of big data, this paper puts forward the design and research of BBS under the background of big data. The research breaks through the traditional BBS system design, adopts new multi-layer web architecture, and the Nginx control mode supports the big data cloud services. It uses Python tools to realize the

rapid expression of interface data, and designs multi topic functional modules to meet the needs of the latest users. Based on the above ideas, this paper introduces in detail the refinement of user requirements, the division of functional modules, data model design, technical architecture design, and important interface design. Through the new design of BBS forum, people can communicate on more platforms, and the communication between users will become faster and more convenient, providing richer data resources for big data analysis.

Keywords

Bulletin Board System, Big Data, Web Framework, Data Source

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

BBS (Bulletin Board System)起源于上世纪七十年代末期,其网站论坛是在上个世纪九十年代初伴随着国际互联网而逐渐流行并迅速得到了发展[1]。国际互联网的诞生使以前只能在本地访问的BBS系统可以被全世界网民任意访问,这就使得BBS变得更加有吸引力[2]。通过BBS论坛大家对共同感兴趣的话题展开讨论、表达自己的观点和看法、汲取知识等[3]。BBS论坛存在的形式和种类很多,按照网站的特性以及服务用户来区分,可分为:校园BBS论坛、商业BBS论坛系统、新闻媒体BBS论坛,等等。按照性质来划分,可分为2种形式:一种是商业的BBS站;另外一种是非商业的BBS站。

然而,近些年BBS的发展相对缓慢,原因归结于BBS技术落后,不能满足用户的要求,新的应用环境得不到满足。早在2000年前,BBS还处于聊天室的功能架构时代,基于TELNET技术实现的BBS界面简陋,操作繁琐,很少用户愿意使用。现在的BBS虽说运用了流行Web结构模式设计[4][5],但不能和大数据时代相结合而缺乏资金支持系统不能满足用户的需求。

关于BBS技术的研究可以从两个方面进行讨论:一是BBS技术,关注BBS这类系统的运营先进性,如早期的Telnet技术,信息的动态交互技术[5][6],架构技术[6][7][8],信息表现技术[9][10]等。二是BBS的分析技术,重心研究系统的信息内容的技术[11],如信息挖掘[12][13]、舆情分析[3]等。BBS的分析技术是针对BBS的信息内容方面进行的研究[14][15],这些研究与现代的大数据技术有着非常紧密的联系。

上述介绍可以看出,在这种大数据技术飞速发展的情况下,BBS这类多主题的数据的优势明显,能够通过大数据分析挖掘其潜藏的价值。因此,本文提出的大数据背景下的BBS论坛设计是非常有意义的。BBS将采用全新的多层Web架构,后台使用Nginx控制访问数据,支持云服务,提高数据访问效率。使用Python工具实现数据的快速表达,以及包含更多主题功能信息分类。

2. 论坛分析与架构设计

2.1. BBS的网络支撑

1) 网络编程

BBS系统必须支持网络功能,论坛系统将以PHP编程语言为工具实现系统的各种网络接口功能。PHP是一种嵌入式语言,它的风格和C语言比较类似,在现在目前市面上大多数网站上得到了应用。PHP综

合了 C 语言、perl 等语言的语法，它并且有自己独特的语言特征。这对于与 Web 开发者来说是一种比较强大的开发工具。

2) 高效运营特性

使用 PHP 作为主要开发语言，其源码完全有较高的开放性，方便开发者的代码重用，编程语法也相对简单，可以以极为简单的语法来完成和其它 WEB 开发语言一样的功能；能够运用面向对象的思想实现编程，提供工作效率。另外，该编程语言也具有较强的移植性，PHP 的语言脚本，不仅应用于 WINDOWS 平台，还可以运行在 UNIX 等平台上。

2.2. 论坛的数据存储

1) 存储系统选型

论坛将采用数据库管理系统作为论坛聊天的数据记录平台。目前数据系统有很多种，较为常见的有甲骨文公司的 ORACLE，而研究的论坛系统可以采用 Mysql 数据库作为其数据存取平台，Mysql 数据库是一种小巧型的数据库，它属于关系型数据库，能够将一些关系型数据进行存放，目前 Mysql 数据库的种种特性使得它成为最受欢迎的数据库系统之一。Mysql 数据库相比于 oracle 数据库来说规模比较小，功能也没有 ORACLE 数据库全，数据库体积也不够大，但是正是因为它的这种特性，才使得它的成本低，访问速度更加快，而且，Mysql 数据库目前的种种功能完全能够支撑一个小的网站的运作，Mysql 也是比较适用于初创型的公司。

2) 存储优势

从式架构的存储结构作为该论坛的网络架构。在这种模式下面，使得数据会更加的安全，如果想要访问数据库，就必须发送一个请求给服务器端，这样当服务器端收到这个请求后，相应的返回给客户端应答数据，从而就完成了一次数据传输。

在 Mysql 数据库中设置需要的密码和相应的用户角色，这样当我们进行的一定的安全设置后，使得数据会保存的更加安全。Mysql 数据库用来存放关系型的数据，那么在一个网站架构设计当中，mysql 数据库就会承担关系型数据的保存任务，那么在关系型数据存放这一环节就会更加的快速。

2.3. 信息发布

Web 服务器是一种 BBS 系统信息处理和发布的核心。在服务器上面，可以配置一些应用服务程序，这样所有网站浏览者都能够登录到网站服务器上面进行数据的浏览。它更像是一个商铺，客户通过进入到商铺里面来选择自己所要的物品。

目前市面上存在着多种类型的 Web 服务器，如 Apache、Nginx、IIS 等，通过对比发现，Apache 服务器相对于其它的服务器，它是出现较早，应用更为广泛，它能够运行在几乎所有的计算机当中。

2.4. BBS 的架构

论坛系统沿袭了传统的结构模式，有客户端、服务器端以及数据存放端的整体架构设计，当一个用户访问 BBS，也就是客户端通过 internet 连接到服务器端，服务器端就会接受这个请求，服务器一般采用的是 apache 服务器，这个时候服务器就会选择性的将数据请求分发到后台真正存储数据的服务器，前提是要我们在服务器上面进行设置，这样做的目的是为了较少服务器的压力。如果不在服务器上面做这些优化，全世界的访客都访问这个网站，前端服务器就会崩溃，导致业务不能正常运转，相反，将这些用户请求很好的转发到后台的服务器之后，就会使网站系统不管在什么情况下都能够正常运转。BBS 论坛的网络架构如下图 1 所示。

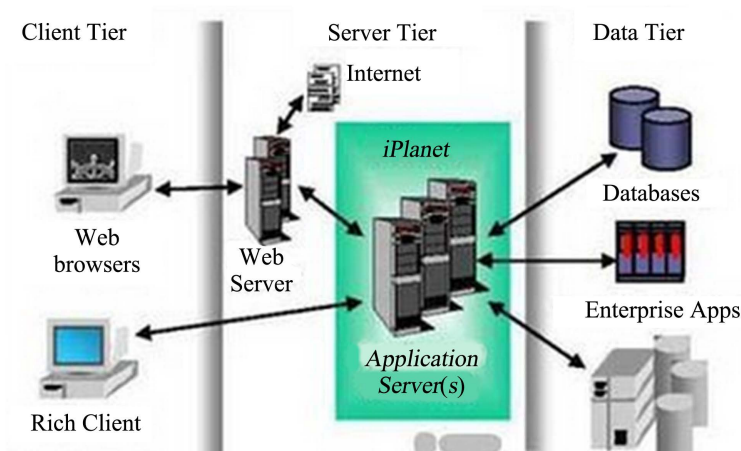


Figure 1. BBS frame
图 1. BBS 架构图

3. 需求与关系

3.1. 需求分析

1) 基本要求

软件的需求分析是根据用户的业务需要进行某些关键点信息的提取的，通过进行关键信息的分析来规划软件产品，这个步骤是软件系统规划的前提，相当于构建一个大的框架，然后通过框架来一步步进行软件产品的开发工作。这一步的成功意味着我们能够开发出一个令客户满意的产品，并且顺利的交付于客户，使得客户能够真正利用我们的软件产品解决业务问题。

为了使得系统的模块能够正常使用、系统更加人性化，以及满足系统的实用性及安全性要求，本系统设置了 4 种角色：游客、普通会员、版主、超级管理员。

2) 主要功能

游客的各项功能为看帖，只有注册了论坛的会员后，才能执行完整的普通会员功能。

普通会员的各项功能有：用户登陆、发帖、看贴、发表评论、修改个人资料等操作。

版主的各项功能：修改本版块、查看版主信息、修改版主信息、查看帖子、管理本版块的帖子、版主拥有普通会员所有功能。

超级管理员的各项功能：用户管理、新闻管理、板块管理、帖子管理、评论管理、添加用户、其他所有功能等。

3) 数据管理

普通会员可以不登录进入系统，系统用户的数据包括用户名、用户密码、用户类型等等。用户 ID 在整个系统中是唯一的。用户浏览帖子时，数据应当包括用户名、帖子主题。用户根据某一主题可以发表帖子，发表帖子需要的数据包括用户名、Email、版块分类名、发表帖子标题、发表帖子内容。用户回复帖子的数据包括用户名、Email、回复帖子序号、回复的内容、版块分类名。管理员进入后台管理时，需要数据，包括管理员 ID、管理员密码。用户只有登录了才能发帖，回帖，管理等等。

3.2. 关系描述

1) 论坛的业务流程

业务流程可以使用用户和功能的关系图来表示。详细见图 2 所示。

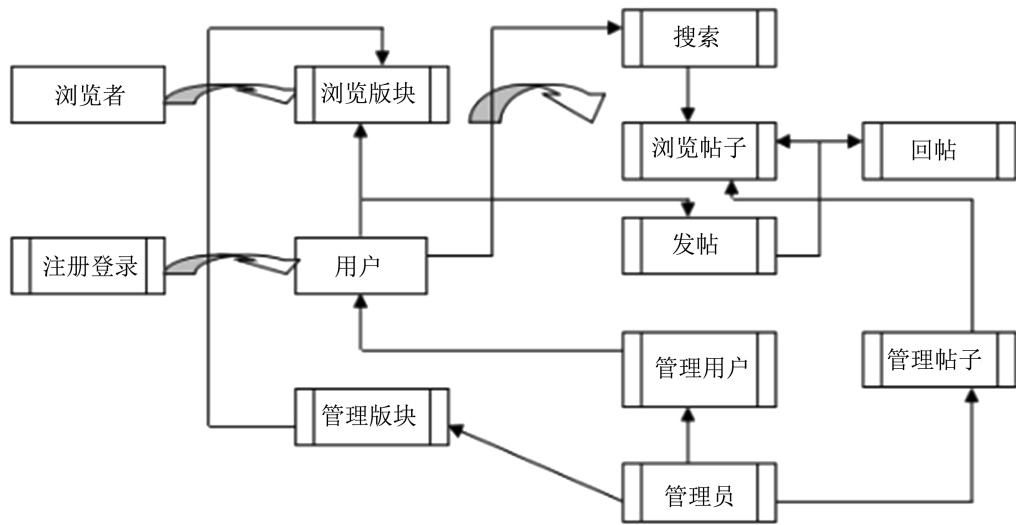


Figure 2. BBS workflow
图 2. 论坛业务流程图

2) 论坛的 ER

BBS 的 ER 关系通过 ER 图的方式给展示了出来。这种关系模型图完全区别于数据库模型 RR 图, ER 图目的在于使得项目所有参与人员都能够直观的理解项目需求, 使相关人员之间相互配合完成项目。ER 图是由不同的实体类型、关系、特性和类型构成。实体是诸如用户的实际对象, 有时更抽象, 但必须有业务意义。特性用于描述实体, 关系用于实体之间:

ER 模型是对现实世界的一种抽象表达方式。主要成分为实体、联系和属性。使用这三种成分, 完全能够建立许多应用于实际环境的 ER 模型。下面贴出本论坛系统所使用的数据库实体, 它们分别为帖子实体、回复贴实体、版面实体、类别实体和用户实体, 如图 3 所示。

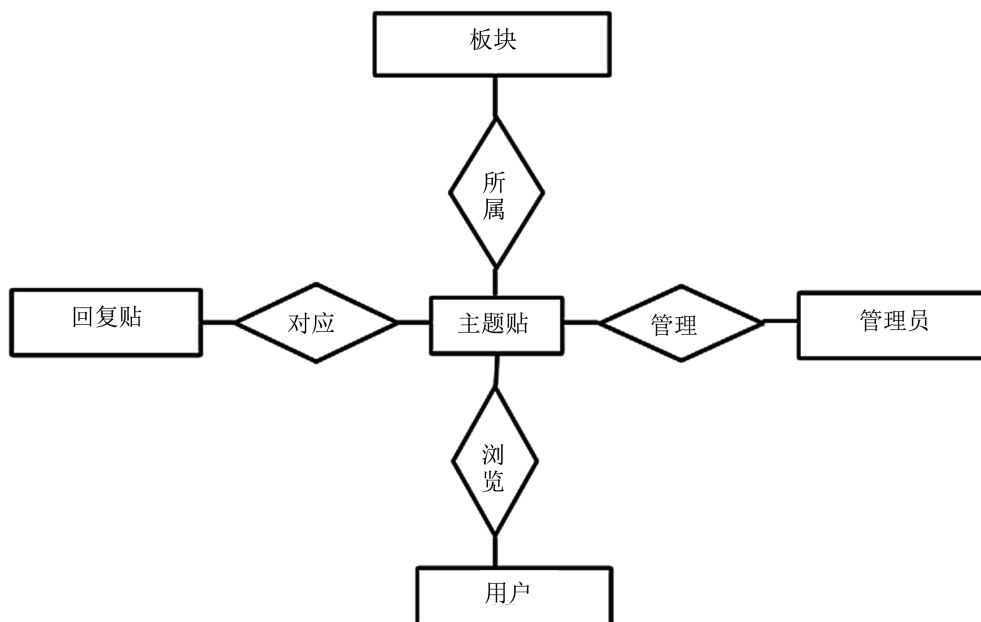


Figure 3. BBS ER model
图 3. 论坛的 ER 模型

3.2.1. 帖子实体

根据实体包括编号、所属版面、标题、内容、发布者、发布时间、表情、对帖子进行操作的时间、是否为置顶帖子、被置顶的时间、是否为精华贴子和被设置为精华贴子的时间等属性。其中是否为置顶帖子与是否为精华贴子属性分别用来标识帖子是否被设置为置顶或是精华贴子，1 表示“是”，0 表示“否”。帖子实体的 ER 图如图 4 所示。

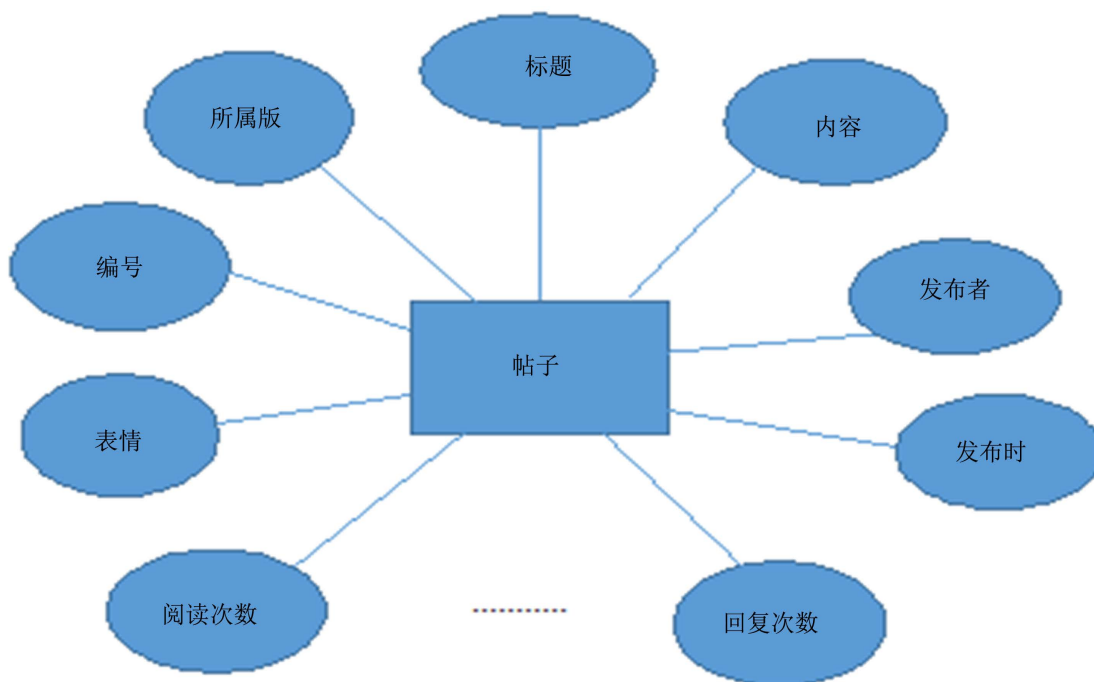


Figure 4. Entry ER graph
图 4. 实体的 ER 图

3.2.2. 回复贴实体

回复贴实体包括了编号、帖子 ID、标题、内容、回复者、回帖时间和表情属性。回复贴实体的 E-R 图如图 5 所示。

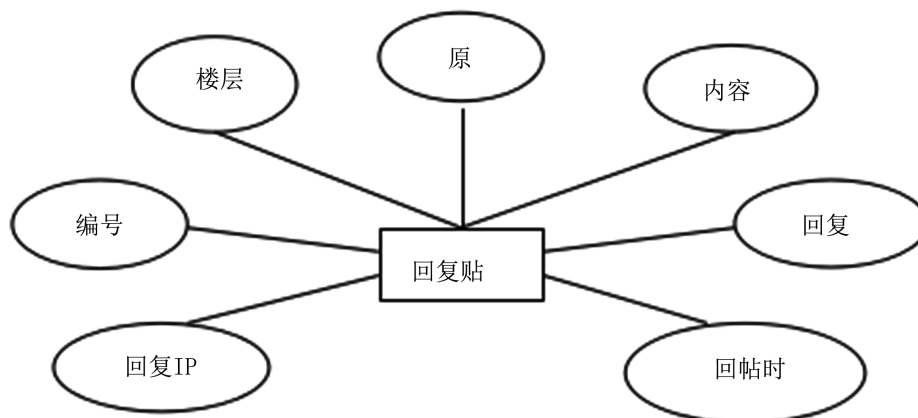


Figure 5. Reply entry graph
图 5. 回复贴实体图

4. 数据模型

4.1. BBS 数据设计

BBS 论坛网站开发，首先要对论坛系统所使用到的数据来进行逻辑抽象设计，不但要使人容易看懂，而且要能够适应 BBS 论坛系统功能的调用，只有这样才不至于导致逻辑上面的不清晰、混乱，与此同时还必须全方位、全方面地考虑数据在什么样的情况下才能保存够使系统调用数据是不会阻塞，使数据库不会成为一个瓶颈，之后，还必须保证数据库的高可用性，不至于使得数据库空间浪费，同时保证数据不会被破坏。

对于一个网站系统或者一个程序的整体开发来说，数据库的选用以及设计显得非常的重要，数据库设计的好坏就预示着程序开发工作后期的便捷以及能够确保工作顺利完成。在一个大型网站开发的过程中，数据库的选择能够确保互联网用户迅速快捷的连接到网站，获取所需，可见数据库开发的重要性。

4.2. 结构表

根据论坛的数据特征，设计了下列 8 种数据模型，用来表达论坛的 8 种数据对象，其对应着 8 种数据表。

如 block 表(block_id, block_name, user_id, block_number, block_bulletin, block_id, block_explain, block_icon); collection 表(collection_id, post_id, user_id)等等。

还有表 count, news, online, post, reply, user 等。表 user 如图 6 所示。

| 字段 | 类型 | 整理 | 属性 | Null | 默认 | 额外 | 操作 |
|---|--------------|-------------------|----|------|------------------------------|----------------|----|
| <input type="checkbox"/> user_id | int(10) | | | 否 | | auto_increment | |
| <input type="checkbox"/> user_name | varchar(10) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_password | varchar(50) | latin1_swedish_ci | | 否 | | | |
| <input type="checkbox"/> user_sex | varchar(2) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_birthday | date | | | 否 | | | |
| <input type="checkbox"/> user_identity | int(1) | | | 否 | 1 | | |
| <input type="checkbox"/> user_realname | varchar(10) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_email | varchar(50) | latin1_swedish_ci | | 否 | | | |
| <input type="checkbox"/> user_regtime | date | | | 否 | | | |
| <input type="checkbox"/> user_avatar | varchar(100) | latin1_swedish_ci | | 否 | user_avatar/avatar.jpg | | |
| <input type="checkbox"/> user_avatar_small | varchar(100) | latin1_swedish_ci | | 否 | user_avatar/avatar_small.gif | | |
| <input type="checkbox"/> user_level | int(3) | | | 否 | 0 | | |
| <input type="checkbox"/> user_signature | varchar(200) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_address | varchar(100) | gb2312_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_zipcode | int(6) | | | 否 | | | |
| <input type="checkbox"/> user_qq | int(10) | | | 否 | | | |
| <input type="checkbox"/> user_zodiac | varchar(2) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_constellation | varchar(2) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_bloodtype | varchar(2) | gbk_chinese_ci | | 否 | A | | |
| <input type="checkbox"/> user_phone | int(11) | | | 否 | | | |
| <input type="checkbox"/> user_profession | varchar(20) | gbk_chinese_ci | | 否 | | | |
| <input type="checkbox"/> user_xueji | varchar(10) | gbk_chinese_ci | | 否 | 0 | | |
| <input type="checkbox"/> user_experience | int(10) | | | 否 | 0 | | |

Figure 6. User table

图 6. User 表

5. 论坛接口设计

论坛根据需要其接口至少包括四类：论坛主页，个人中心，论坛聊天室，论坛管理等。其接口采用 PHP 编程实现，容器中的数据来源于后台数据库，通过网站 Apache 控制数据的访问。

如个人中心，如下图 7 所示。

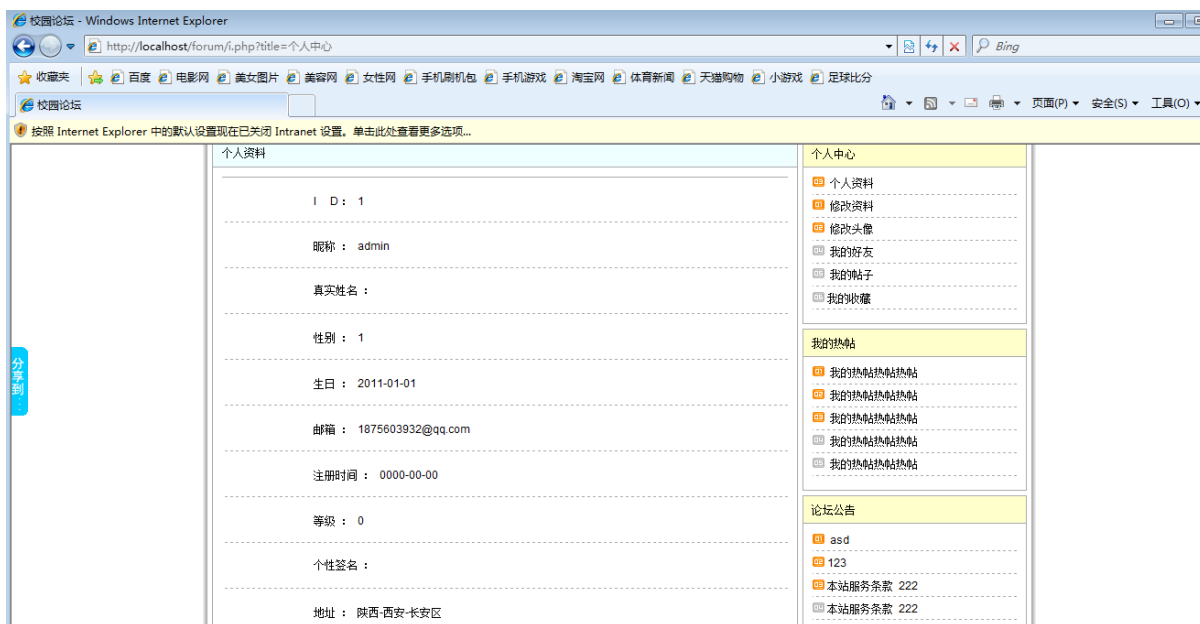


Figure 7. Person center

图 7. 个人中心

论坛聊天室如下图 8 所示。



Figure 8. Chat room diagram

图 8. 聊天室图

6. 结论

1) 新的 BBS 设计为大数据分析开辟了新的数据源。BBS 信息交流系统主题范围广、信息量大，是大数据分析非常好的来源，能为大数据提供各种有价值的数据分析。

2) 进一步拓展了 BBS 论坛的 Web 技术应用。使用现代 Python 作为制造工具界面友好, 能吸引更多的网络用户。更加细致深入地理清了 BBS 的功能属性, 为 BBS 的发展注入了活力。BBS 实时互动的新模式对网络运营商有着较强的吸引力。

3) 有着更为广泛的论坛分类、支持多种形式和数据类型, 以及多种辅助功能, 这将是 BBS 未来发展的方向。例如天气预报、广告、游戏以及其他的一些生活中的帮助功能。系统的交互性、健壮性还有待增强。

基金项目

海南省科技厅 2022 年自然科学基金重大专项(高层次人才项目)《拓扑向量空间辨识及其多网络车辆智能诊断应用》; 海南省科技厅 2020 年自然科学基金重点项目《复杂网络理论特性及其优化应用研究》(60MS060); 教育部 2020 年项目《新工科研究与实践项目》(E-JSJRJ20201333); 海南省高等学校 2020 年教育教改项目《在线课程智慧平台的构建及学习状态演化分析研究》(Hnjg2020-120)。

参考文献

- [1] 田贤忠, 姚明超, 顾思义. 基于 BBS-LDA 的论坛主题挖掘[J]. 浙江工业大学学报, 2020, 48(1): 55-62.
- [2] 丁晓阳, 王兰成. 网络论坛文本特征词权重计算优化方法研究[J]. 情报理论与实践, 2021, 44(5): 187-192.
- [3] 李建新. 基于数据流 MSW 算法在 BBS 舆情分析系统中的应用[J]. 合肥工业大学学报(自然科学版), 2019, 42(2): 195-199.
- [4] 苏恒阳. BBS 热点引导技术在动态 Web 上的实现[J]. 科技通报, 2014, 30(4): 155-157.
- [5] 崔立波. 基于 ASP 的校园 BBS 的设计与实现[J]. 电子技术与软件工程, 2015(21): 49-50.
- [6] 毋世晓. 基于 MVC 的 BBS 论坛创建与实现[J]. 电脑编程技巧与维护, 2019(12): 21-23.
- [7] 杨道全, 许勇, 张凌. CNGI 高校联盟交流平台中 BBS 系统的设计与实现[J]. 广西大学学报(自然科学版), 2011, 36(A01): 276-281.
- [8] 何博文, 郑剑. BBS 论坛系统设计与实现[J]. 电脑知识与技术, 2018, 14(18): 42-44.
- [9] 高云峰. 对基于 Java 的 BBS 论坛系统的设计与实现分析[J]. 电子技术与软件工程, 2018(4): 202.
- [10] 黄俊霞. 基于 NET 的高职教师学术论坛系统设计与实现[J]. 湖北第二师范学院学报, 2017, 34(2): 41-47.
- [11] 熊熊, 罗春春, 张焯. 股吧和交易: 股吧中的信息内容研究[J]. 系统科学与数学, 2017, 37(12): 2359-2374.
- [12] 涂志芳. 联结的力量: 高校图书馆用户联结机制研究——以未名 BBS 图书馆版为例[J]. 大学图书馆学报, 2019, 37(1): 87-93.
- [13] 曹丽娜, 唐锡晋. BBS 话题的地理分布分析[J]. 系统科学与数学, 2016, 36(5): 671-682.
- [14] 王丽英. 高校 BBS 教育舆情的时空特征模型构建和热点发现——基于南京大学小百合 BBS 数据集的分析[J]. 现代情报, 2016, 36(1): 84-89.
- [15] 李永智, 耿绍宁. 建设高校网络新平台新高地——高校 BBS20 年回顾[J]. 中国高等教育, 2016(19): 43-46.