电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE

论文题目 PHPMyFAQ 系统的设计与实现

专业学位类别	工 程 硕 士
学 号	201192230436
作者姓名	于 冰
指导教师	吴 跃 教 授

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：_____ 日期：2015年 5月 10日

论文使用授权

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后应遵守此规定）

作者签名：_____ 导师签名：_____

日期：2015年 5月 10日

分类号 _____ 密级 _____

UDC [注1] _____

学 位 论 文

PHPMyFAQ 系统的设计与实现

（题名和副题名）

于 冰

（作者姓名）

指导教师 吴 跃 教 授

电子科技大学 成 都

丁项平 高 工

山西省定襄县昌得利机械加工贸易有限公司 忻州

（姓名、职称、单位名称）

申请学位级别 硕士 专业学位类别 工程硕士

工程领域名称 软 件 工 程

提交论文日期 2015.3.25 论文答辩日期 2015.5.10

学位授予单位和日期 电子科技大学 2015 年 6 月 22 日

答辩委员会主席 _____

评阅人 _____

注 1：注明《国际十进分类法 UDC》的类号。

THE DESIGN AND IMPLEMENT OF THE PHPMYFAQ SYSTEM

A Thesis Submitted to

University of Electronic Science and Technology of China

Major: **Master of Engineering**

Author: **YuBing**

Advisor: **WuYue**

School : **School of Information and Software Engineering**

摘要

随着Internet技术的广泛应用和Web技术的不断发展，对传统的答疑方式产生了深远的影响。基于网络的答疑系统成为当今计算机应用的一个热点。采用非面对面教学方式的教学辅助系统，学生提问和教师答疑可以不受时间和地点的限制，同时也避免了老师大量重复回答问题。体现了网上答疑的开放性、交互性和自主性等特点。

PHPMyFAQ 是本论文研究用于辅助教学的自动问答系统。分析三层架构模式体系结构。根据三层架构的优点和对网络答疑系统的分析，网络答疑系统更适合采用三层架构模式。研究对比动态网页的技术，并介绍系统开发环境，系统开发采用 Apache+PHP+MySQL 的黄金组合，并分别介绍其优点以及网络数据库的构架。利用的中文分词的方法、特点及难点进行归纳说明，然后对文本自动答疑中所采用的正向最大匹配法进行阐述，并给出算法和对歧义解决的方法等。对系统所要完成的功能进行详细的分析设计以及实现，并就实现过程中涉及的关键技术予以重点阐述。分析和总体设计，将面向对象的分析和设计技术应用于系统的设计，引入软件工程的Case管理方法，利用Rational Rose 2005这个辅助工具，获得良好的系统解决方案，开发出的总体结构，包括用户管理系统、自动答疑系统、异步和同步答疑系统。它们相辅相成来满足系统的总体需求，然后给出系统的功能模块的设计、数据库设计及系统各部分功能的组成。最后进行了系统的测试，包括压力测试、安全测试、浏览器兼容性测试，证明系统能够正常运行。

关键词：自动问答系统，三层架构，教学辅助系统

Abstract

The popularization of the Internet and fast development of Web technology have significant impact on traditional MISs. The net answer system is one of hot areas of computer applications, asking and answering questions between students and teachers are not restricted by the space-time in the network teaching a teaching form being not face to face. It represents the characteristics of long-distance education, such as opening, interacting, self-determining and so on.

The research of this thesis can be used on the teaching of automatic interrogator-responder system, the b/s three-tier architecture model was applied in solution, advantages of this three-layer architecture are discussed. The platform uses the technology of the advanced active server pages. This thesis uses php network programming language to develop the system. The chinese word segmentation using javacc and ictclas perfects the chinese searching functionThis paper is based on the maximum matching method, and takes the context into account. Chinese word is matched to dictionary in memory by binary search. This is very efficient in precision and saving time. To obtain good system solutions, puts forward the overall structure of question answering system, including user management system, automatic question answering system, question answering system and synchronous asynchronous answering system. They complement each other to meet the overall system requirements, then each function part design and system function module design, the database system are given. Finally, the system testing, including testing, security testing, pressure browser compatibility testing, the system can run normally.

Keywords: teaching of automatic interrogator-responder system; three-tier model; net answer system

目录

第一章 前言

1.1 论文选题的背景及来源

随着网络技术的迅猛发展，传统的教学手段必须引入新的模式，才能够满足当前海量信息的教学内容需求，为适应当今社会发展对高素质创新型人才的需要，创造一个在教师指导下的学生自主式学习环境，是当前的研究热点之一。网络教育形式融合了信息技术与教育，使得学习机会更便捷，学习环境更多样，学习资源更丰富，学习活动更自主、个性化，增强了教育的适应性，为教育注入了新的活力，因此近年来备受关注。

人们通过网络教育这种新的教育方式，可以不受时间和空间限制，展开以学习者为中心的学习，但这种教育方式也由于发展历程较短而存在不足，师生之间的交流弱化是很重要的一点。“师者，所以传道授业解惑也”，传统教育的“解惑”建立的基础大多是面对面的交流，而通过网络的交流不再直接、形象，阻隔了解惑的道路。于是，网络教育中一个值得关注的重要环节就是，如何为参加网络学习的学生疏通解惑渠道，给他们的答疑需求以满意答复。答疑系统要高效、科学，不仅可以快速响应学生、及时解答疑难、消除学习障碍，也可以为教师减轻工作压力。总之，答疑问题的解决是有效发挥远程教育优势的关键。

在当前的教育领域中，发展基于 Web 的远程教育的核心问题之一就是研究与开发答疑系统，也就是怎样在基于 Web 的自主式学习环境中实现有效答疑。作为远程教育系统的重要组成部分之一，答疑系统决定了整个系统的效果和效率，受到越来越多的教育者和远程教育系统开发者的关注。基于 Web、高效、专为教学服务的远程答疑系统的建立已成为现代教育和计算机应用研究的一个重要领域[1]。在人工答疑之外，学生所提的一部分疑问能被计算机自动解答，将使得教师负担大大减轻，同时也使得学生在学习过程中所遇难题能被及时排解，学生网络学习的效率和质量因而得到提高。但是，高效、准确、智能化的自动答疑系统的开发在当前仍是一个技术难题。针对这一要求，本论文提出利用现有的信息检索和 XML 技术，对现有的智能模式进行改革，对网络资源进行充分利用，由机器处理、自动理解以自然语言描述的用户提问，人工或自动的返回相应答案，达到为教师减负、为学习助力、提高网络教学质量和效率的目的，因此本论文的选题十分有意义。

1.2 论文研究的目的及意义

本文针对目前答疑系统的现状，将面向对象的分析和设计技术应用于系统的设计，引入软件工程的Case管理方法，利用Rational Rose 2005工具，获得良好的系统解决方案，提出的总体结构，包括用户管理系统、自动答疑系统、异步和同步答疑系统，解决了现在答疑系统的智能化及快速响应用户请求的要求。这也是本研究的目的。

汉语自然语言处理和文本检索等技术是网上自动答疑系统的核心技术，这些技术也常被用在网上搜索引擎中，网上搜索质量的提高很大程度上来源于这些技术的发展。答疑不仅能帮助学生巩固系统学习的内容，同时也能对课堂教授内容进行有益补充。因此，无论采取何种学习形式，答疑始终是学习活动的必要环节，不可或缺。由此可见，网络为答疑引入了新的内涵，基于网络答疑系统的研究和开发也就具有了较高的实践价值和教育意义[2,3]。本论文将把自然语言处理技术和网络技术等结合并加以应用，从而建构一种网络化的基于本体论的智能课程答疑系统。这对于提高答疑效率和质量、扩展答疑空间，从而提高教学质量，都有着非常重要的意义。

1.3 国内外研究现状

目前，随着网络技术的飞速发展和应用的逐步深入，基于网络的在线答疑已经是比较常见的事情，我国最初的在线答疑系统源于上世纪1990年代的电脑知识答疑[1]。答疑的内容随着信息技术和网络的发展已经有了较大的变化[2]。较为流行的网络答疑中，目前有各类网络公司的答疑，包括Microsoft的认证答疑，以及各类远程教育机构所组织的答疑等[32]。在答疑系统的产品上，比较具有代表性的有如下几类[4]：北师大开发出的网络在线答疑信息系统，基于 web 技术进行开发，能够对所有参与答疑人员的信息进行统一管理和操作设置，通过一个独立机构对这些信息进行维护，以保证答疑的客观性以及考生信息的保密性；并且接受社会各界的开放使用，进行基于网络的题目征集和答疑信息发布，并组织答疑，进行成绩分析等。美国ALA公司中国分部所开发的答疑内容服务信息系统，则是为各个实体提供在线答疑服务的产品[5]。该软件能够基于局域网或者广域网来组织各类答疑，还能够支持考生进行在线的模拟练习等[6]。同时为授课教师提供入口，与学习者进行线上的问答交流，进行知识的辅导等[7]。我国国家信息中心结合当前的中小学教育模式，也开发出了针对性很强的软件产品，支持注册用户进行在线学习、互动练习以及交流讨论，也支持教师用户在网上进行在线辅导和作

业指导等[8]。紫晶远程答疑系是加拿大奥兹公司所开发的培训答疑信息系统，该系统支持用户进行基于各类职业的培训与答疑，并结合一些权威证书的要求对学员进行针对性的培训[9]。而基于分布于在线的专用答疑信息系统目前还不多见，本文将对在线答疑系统进行需求分析和详细设计。

1.4 本文研究的主要内容与组织

FAQ 系统的国内外研究现状表明：网络教学系统需要在设计和实施时考虑的重要问题之一是，如何优化网络教学模型以最大化利用互联网的优势，及时有效地解答参与网络教学的学生疑问，同时为参与教师节省时间、提高效率，并且能重复利用问题与解答这一资源。网络教学中另一重要课题是，如何使师生之间通过网络实现更方便、更有效的交互。教学工作中，教师向学生答疑是一个重要环节。学生能够通过异步、同步或自动形式的答疑巩固所学知识，教师也可以通过回答学生问题来了解学生学习中存在的难点，从而对教学内容和方式进行调整改进。另一方面，学生还可以通过浏览其他人提出的问题和解答，吸取经验教训，在今后的学习过程中少走弯路。在答疑库中将来自学生们的问题和来自教师的相应解答有机的组织起来，利用自然语言理解技术，自动分析并匹配学生提出的问题，从而自动给予解答。学生可以很快得到解答，教师也不会有太大的工作压力。教师可以主动将一些有价值的问题预先放到答疑库中去，从而大大节省了讨论时间，将主要精力放在重点和难点上。答疑系统是组成网络教学平台必不可少的部分，它在增进教师和学生间的交流以及帮助学生发现问题并获取答案方面具有重要作用。结合前人在网络答疑系统的研究成果，采用了 PHP 技术（Hypertext Preprocessor，超文本预处理语言）分析了 PHPMyFAQ 系统。针对面对面答疑的不足之处，本系统利用互联网的优点，使得学生和教师之间的答疑可以通过互联网进行。这种答疑方式更加灵活，不受时间和空间的限制，效果更加显著。系统提供的页面包括：学生按章节浏览问题、学生提出问题、教师按章节浏览已回复问题和未回复问题、教师回复问题、教师查看、新增、修改、删除数据库中数据表的内容。系统实现了显示数据、保存数据、修改数据、删除数据功能。网络包括用户管理模块、自动答疑模块、异步答疑模块和同步答疑四个模块。而其中的自动答疑模块就是即将研究的PHPMyFAQ系统。后续论文由六部分组成，其内容组织如下：

第二章分析三层架构模式体系结构。根据三层架构的优点和对网络答疑系统的分析，网络更适合采用三层架构模式。研究对比动态网页的技术，并介绍系统开发环境，系统开发采用 Apache+PHP+MySQL 的黄金组合，并分别介绍其优点

以及网络数据库的构架。

第三章介绍答疑系统分析和总体设计，将面向对象的分析和设计技术应用于系统的设计，引入软件工程的Case管理方法，利用Rational Rose 2005这个辅助工具，获得良好的系统解决方案，提出答疑系统的总体结构，包括用户管理系统、自动答疑系统、异步答疑系统和同步答疑系统。它们相辅相成来满足系统的总体需求，然后给出系统的功能模块的设计、数据库设计及系统各部分功能的组成。

第四章主要对是PHPMyFAQ系统的学习与研究，对系统所要完成的功能进行详细的分析设计以及实现，并就实现过程中涉及的关键技术予以重点阐述。

第五章对匹配模块利用的中文分词的方法、特点及难点进行归纳说明，然后对文本自动答疑中所采用的正向最大匹配法进行阐述，并给出算法和对歧义解决的方法等。本论文采用的中文分词方法主要应用在自动答疑系统中对中文文本的处理。

第六章是系统的测试以及效果演示部分。

第七章是结论。在总结主要特点的同时，指出其中的不足之处。

第二章 答疑系统的相关理论与技术

2.1 PHP 技术

PHP（Hypertext Preprocessor：超文本预处理程序）是一种当今 Internet 上较为成熟的服务器端脚本语言，它是专门为 Web 设计的。工作原理图如图 2-1 所示:

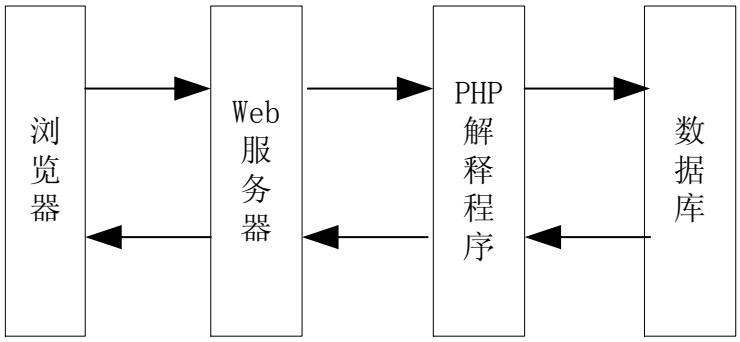


图 2-1 PHP 工作原理图

PHP 技术在本文 2. 2 小节做了简要介绍，本小节就其主要特点做出归纳总结，PHP 在很大程度上综合了 Perl， Java 和 C 语言的精华，在语法架构上继承了 C 语言的风格，可以比 CGI 或者 Perl 更快速地执行动态网页。图 2-2 所示为 PHP 工作流程示意图：

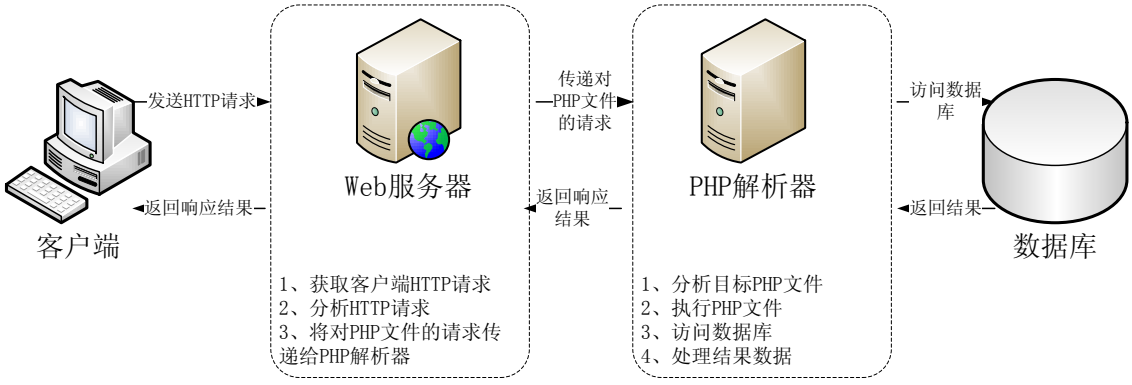


图 2-2 PHP 工作流程图

PHP 不仅易学易用，还具备各式各样的强大功能，并可以根据需要扩展这些功能。PHP 在定义了良好的应用编程接口（API）的同时，也提供了丰富的函数集。跨平台可移植性是 PHP 的良好特性，能够正常运行于基于 Windows、Unix 和 Linux 系统的 Web 服务器上，支持多种流行 Web 服务器（包括 IIS、Apache 等在内）。PHP 引擎优化了 Web 应用程序的响应时间，在一些情况下成为了 Web 服务器本身的一部分。另外，当工作平台发生变换时，用户不需要更改 PHP 代码，可

以直接在新的平台上使用。在 PHP 下，从客户端来的每一个信息资源都被作为出现在 PHP 脚本中的一个变量，处理方式相同。PHP 是一种服务器端进行解释的语言，由服务器返回客户端的是 HTML 页面，因而安全性极高。PHP 针对 Web 应用设计，所以很方便完成简单的数据库连接和查询。PHP 体系结构如图 2-3 所示：

PHP 不仅易学易用，还具备各式各样的强大功能，并可以根据需要扩展这些功能。PHP 在定义了良好的应用编程接口（API）的同时，也提供了丰富的函数集。跨平台可移植性是 PHP 的良好特性，在基于 Windows、Unix 和 Linux 系统的 Web
平台上都可以应用。

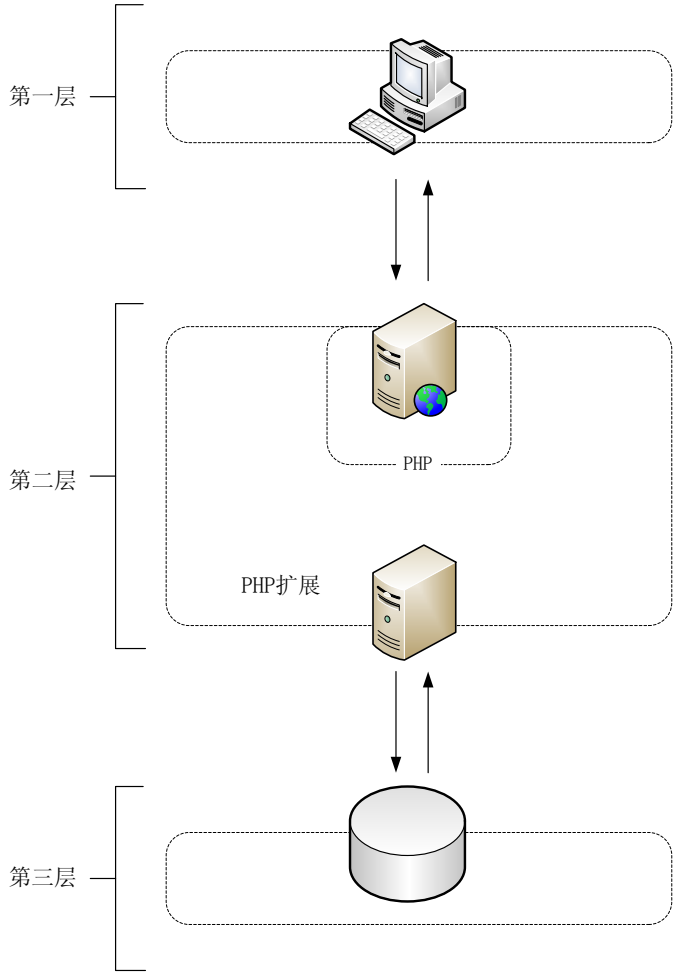


图 2-3 PHP 体系结构

PHP5. 0 具有如下特点[6-8]：

- 开放源代码，零运行费用。
- 基于服务器端，即使 PHP 程序很大、很复杂，也不会影响客户端的运行速度，激活 Activex 控件很方便，可动态生成 Script 语句。

- 跨平台特性强，在不同平台间移植无须修改代码，语言简单，还可以嵌入到 HTML 内部，简单易学，编程灵活。
- 模块化设计，效率高，扩展性强，消耗系统资源少。
- 用户可以组建 PHP 新版本，可以读取 XML 信息。
- 高效支持数据库多达十几种。
- 支持生成和处理动态图像。
- 有许多支持文件存取、处理字符串的函数，其中包括模式匹配的能力，为标量、数组、关联数组等变量提供支持，可以利用引用来赋值变量，支持布尔数据类型，扩展了 API 模块，即接口模块。
- 资源释放自动化，提供十分智能和通用的生成进程，支持 COM/DCOM，可以对 COM 对象进行无限访问和存取。
- 支持面向对象，设计了良好的面向对象特性。
- 加密机制完整，提供 FTP 支持，可作为个性化模块成为 IIS 插件。具有很好的数据库支持能力，为各种数据库系统提供用法相似的函数，使用更方便。提供丰富的数据库函数，可以方便地对 Sybase、Oracle、MySQL、Solid、InterBase、ODBC 等类型的数据库进行存取。

2.2 动态网页技术

在基于 PHP 的网上答疑系统中，无论是搜索、查看、提出问题，还是参与讨论，用户与服务器端数据库交互都要通过客户端浏览器。动态网页是为了实现交互而需要用到的一种技术。该技术是指网页的内容更新与显示是动态的。动态网页技术不同于传统静态网页技术的地方在于，根据访问者的需求不同，它对访问者输入信息做出的响应和提供响应的信息也不同。图 2-4 说明了执行动态网页请求的过程。

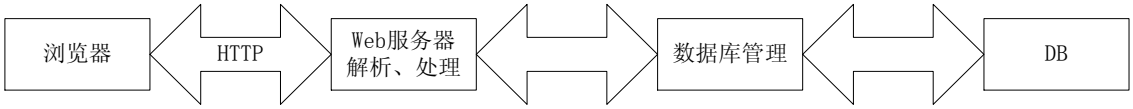


图 2-4 执行动态网页请求示意图

由图 2-4 可知，实现动态网页技术的原理是：在 Web 服务器中保存动态页面，这些页面使用了不同的技术编写，当对动态页面的访问请求从客户端用户向 Web 服务器发出时，用户所访问页面的后缀名将被 Web 服务器用来确定该页面所使用的网络编程技术，然后提交该页面给相应的解释引擎；通过扫描整个页面，解释引擎找到特定的定界符，并将位于定界符内的脚本代码执行以实现不同的功能，

如发送电子邮件、访问数据库、执行算数运算或者逻辑运算等等，最后给 Web 服务器返回执行的结果；最终，Web 服务器传送给客户端的包括解释引擎的执行结果、构成页面的 HTML 内容以及各种客户端的脚本。尽管用户在客户端接收到的页面的表现形式和传统页面是一样的，但实际情况是，该页面的内容是经过服务器端处理而动态生成的，可以进行个性化的设置[5]。

本文主要通过 PHP 技术实现动态网页。PHP 是一种可以嵌入 HTML 文件的脚本语言，其语法借鉴了 C、JAVA 和 PERL 语言，并具备自己的独有风格，能帮助 Web 程序员实现动态网页的快速开发。作为一种脚本语言，它在当今 INTERNET 上使用最广，不需要太多的编程知识就能建立一个用于交互的 Web 站点。本文 2.3.2 小节对 PHP 做了更详细的说明。Apache 是当今应用范围最广的 Web 服务器软件之一，能够运行在 Windows 9X/NT/2000/7 、Unix、Linux 等许多平台系统之上。据调查资料显示，全互联网的 Web 服务器超过半数都是用 Apache 搭建的，近乎所有与网络管理相关的教程中关于 Web 服务内容的讨论都包括 Apache。Apache 的模块化体系结构如图 2-5 所示.

正是由于 Apache 本身具备很多优异的性能，所以成为了很多用户的选择。这些性能表现在以下的一些主要方面：

● 完全免费。作为源代码开放的自由软件之一，不仅编译好的 Apache 程序可以被人们所使用，而且 Apache 的源代码也可以被获取。这一点在用户搭建个人服务器的时候很重要。

● 灵活性和可扩展性。Apache 自己称自己为“补丁服务器”，该说法实际上意味着模块化是它的一大特性，利用这一特性，开发者能够通过添加第三方功能模块从而容易的添加任意功能。Apache 主程序处理流程如图 2-6 和 2-7 所示：

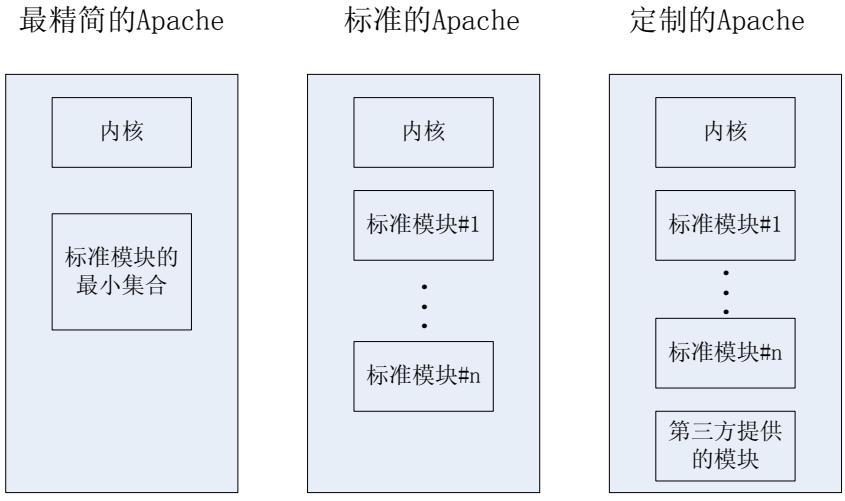


图 2-5 Apache 的模块化体系结构

Apache 最初是针对 Unix 和 Linux 系统开发的，发展了几年之后，Apache 现在已经可以运行在几乎所有的、包括最常用的 Windows 的系统平台下。Apache 服务器软件虽然不是最快的，但其稳定性高却是负有盛名。Apache 的安全性在所有的 Web 服务器软件中是首屈一指的，很少被发现漏洞，而且一旦发现，很快就会有相应的补丁程序被开发出来，因而安全性得到了很好的保证。

2.3 MySQL 技术

数据库的访问是交互网站的关键技术，基于 Web 数据库的应用系统分为三层：客户端是第一层，Web 服务器是第二层，数据库服务器是第三层。本系统采用 MySQL 建立数据库，通过 PHP 在网上进行查询、更新等数据库操作。MySQL 数据库系统以其免费、跨平台、源代码开放、访问效率高、使用方便、独特的权限系统等优秀特点，成为建设动态 Web 站点的主流数据库。MySQL 与 PHP 都可在 Unix、Linux 和 Windows 等流行的操作系统下运行，所以在需要时，基于 MySQL 和 PHP 的程序可以轻易地在不同的系统平台间移植。因此面对互联网的快速发展，PHP+MySQL 是进行 Web 开发的最佳选择。

MySQL 的主要优势是[9, 10]：

- 稳定，速度快，易于使用，价格较低，支持查询语言。
- 安全性和连接性好，MySQL 是完全网络化的数据库，其可以在互联网的任何地方访问，并进行访问控制。
- 性能优越，服务器可与多台客户机同时相连，多个数据库可被多个客户机同时使用，可通过提供多种查询输入方式与结果回显的界面对 MySQL 进行交互式访问。
- 可移植性强，可以在不同版本的 Unix 系统中运行使用 MySQL，同样也可以在不同版本的 Windows 系统中使用。

通过 Web 方式查询数据库的基本步骤是：

1) 对来自用户的数据进行检查并过滤。
2) 与数据库建立连接。
3) 向数据库发起查询请求。
4) 从数据库获得查询结果。
5) 据数据库的查询结果向用户返回相应的内容。
6) 释放与数据库之间的连接。

由于 Apache、PHP、MySQL 的跨平台特性都很好，几乎可以运行在所有的操作系统平台上。将这种技术平台作为搭建基础，使本系统得到了很大的实施弹

性，用户可以选择在不同性能系统平台上按自己的喜好和用户访问量实施。移植到其它的操作系统平台时，只须做很少量的修改。

2.4 UML 简介

UML 的定义包括 UML 表示法和 UML 语义两个部分[11]。

1) UML 表示法定义 UML 符号的表示法，为开发者使用这些文本和图形符号语法提供了标准。这些文字和图形符号所表达的是应用级的模型，从语义来看它是 UML 元模型的实例。可以用五类图（10 种图形）来定义统一建模语言 UML：用例图（Use case Diagram）、类图（Class Diagram）、对象图（Object Diagram）、包图（Package Diagram）、活动图（Active Diagram）、协作图（Collaboration Diagram）、顺序图（Sequence Diagram）、状态图（State Diagram）、构件图（Component Diagram）与部署图（Deployment Diagram）。有机地结合这些图，就可以对系统进行一致的分析与构造，而且这些图还提供了多种系统分析开发的图形表示。

2) UML 语义定义了基于 UML 的精确元模型。元模型对 UML 的所有元素提供的定义性说明在语法和语义上简单、一致、通用，使开发者能在语义上取得一致。UML 还支持对原模型的扩展定义。

当系统采用面向对象技术进行设计时，第一步是描述用户的需求；第二步是构造系统的结构，即根据需求建立系统的静态模型；第三步是描述系统的行为。

在上述前两个步骤中都是建立的静态模型，包括类图（包含包）、对象图、用例图、组件图和部署图等五个图形，属于统一建模语言 UML 的静态建模机制。

其中用例图的典型模式为：

图 2-8 用例图的典型模式

用例图之间的包含关系为：

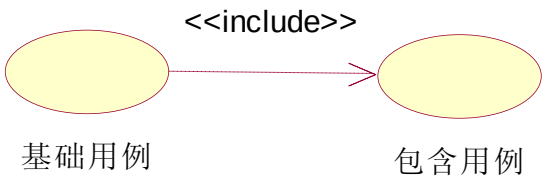


图 2-9 用例图之间的包含关系

用例图之间的扩展关系为：

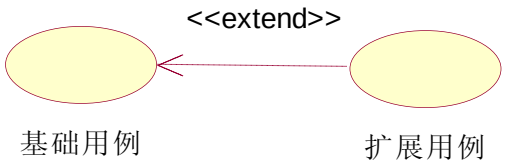


图 2-10 用例图之间的扩展关系

用例图之间的泛化关系为：

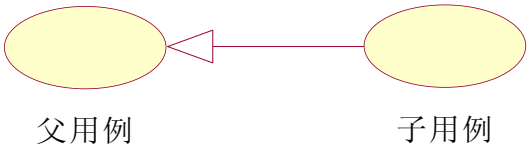


图 2-11 用例图之间的泛化关系

类与类图的典型模式为：

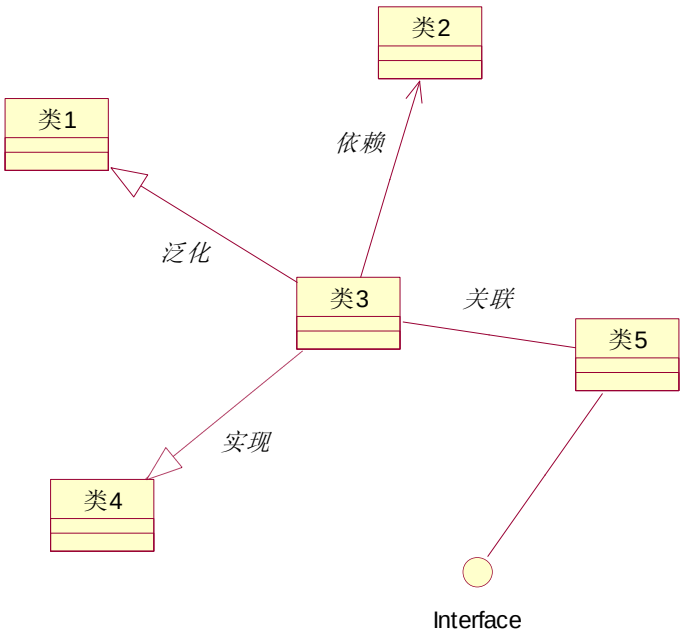


图 2-12 类与类图的典型模式

顺序图的典型模式为：

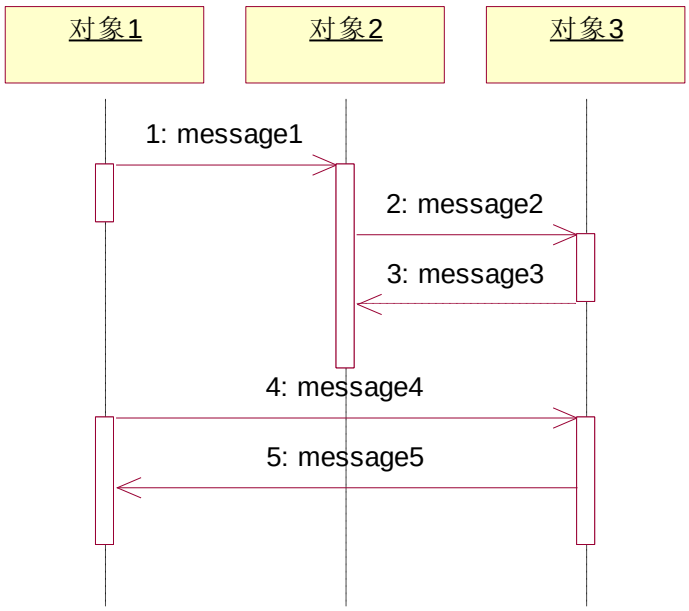


图 2-13 顺序图的模式

活动图的典型模式为图 2-14：

组件图的典型模式为图 2-15

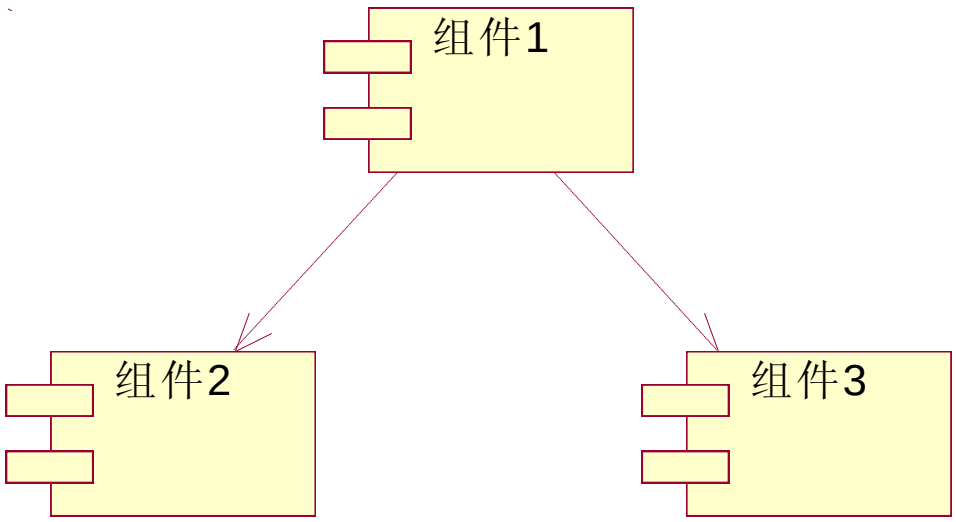


图 2-15 组件图的典型模式

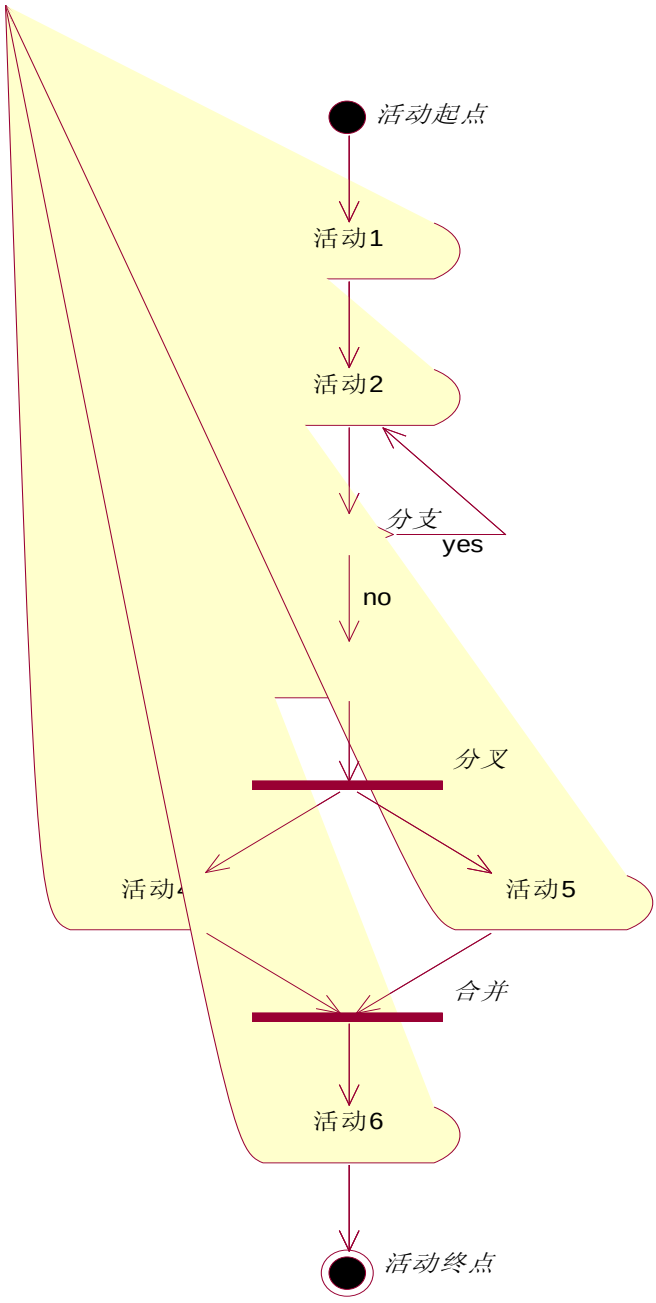


图 2-14 活动图的典型模式

配置图的典型模式为：

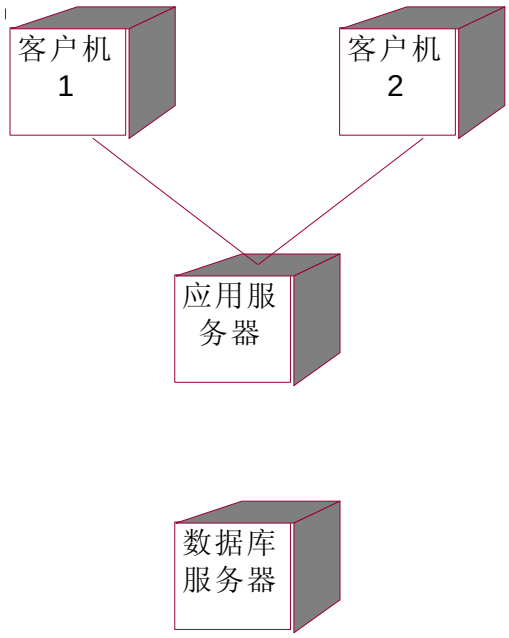


图 2-16 配置图的典型模式

而第三步中建立的模型或者可以执行，或者表示执行时的时序状态或交互关系。它包括状态图、活动图、顺序图和合作图等四个图形，是标准建模语言 UML 的动态建模机制。因此标准建模语言 UML 的主要内容可归纳为静态建模机制和动态建模机制两大类。

UML 是对软件系统的制品进行构建、详述、文档化、形象化的一种语言。面对日益庞大和复杂化的系统，模型和清晰性已经成为它的标志性特征。UML 也已成为一种被完美定义的标准，被广泛接受和响应。因此采用 UML 组建面向对象的系统和基于组件的系统。

2.5 本章小结

本章分析现有三层架构模式以及他们的优缺点，根据网络答疑系统自身的要求，确定了选用三层架构模式作为该答疑系统的结构模式。介绍实现动态网页可能用到的相关技术，并根据本系统的需求，选用 PHP 技术进行开发。介绍系统开发环境相关的 Apache 服务器、PHP 技术及其优点、以及 Web 数据库 MySQL 技术的相关内容。介绍 Web 数据库的构架，分析 PHP 访问数据库的工作流程。

第三章 答疑系统的需求分析与概要设计

学生对于答疑环节的基本要求是及时、有效的获得解答，智能型的网络学习环境根据学生的提问能够将具有针对性的解答进行自动反馈，能够在第一时间通过迅捷的网络传输将反馈送达给学生，从而保证了答疑的有效性和及时性；同时，网络的开放性和超越时空性也将更加多的特色引入到答疑中来。网络答疑转移了答疑的场所——从固定的教室延伸到了无所不在的网络，变换了答疑的时间——从有限的固定时段扩展到了自由随意的时刻，使更多的学生能够参与到该过程中来，同伴之间可以相互答疑，既实现了互动学习，又节省了教师在此中所花费的时间和精力。

3.1 系统的需求分析

下面将通过在业务建模活动中收集的信息来定义所建立的系统，导出待开发系统的所有需求说明，以确保用户的需求得到充分的理解。可以用 UML 的用例视图对客户的需求进行可视化建模。通过用例模型的建立，可以清楚的显示出外部的角色以及它们各自对系统的功能期望。

3.1.1 用例的获取

对系统边界的正确划分是获取用例的首要任务，同时找出涉及的参与者。参与者可以是某种角色（人或事物）的代表，与系统有交互。参与者可以通过分析业务系统找到，结果如下：匿名的一般用户、注册用户、教师以及系统管理员。它们各自的任务分别是：匿名用户可以注册帐户，可以查询问题，可以浏览已存在的问题，但没有资格参加在线讨论；注册用户需要先登录系统，然后可以浏览静态页面，可以输入关键词查询和检索问题，可以查看自己提出的和已存在的问题，并有资格参加在线讨论；解答问题由教师负责，教师还可以查看及修正自己的解答，可以删除用户等等；系统管理员是本系统的后台操作角色，可以维护用户的信息。在分析得出参与者之后，可通过对每个参与者提问题的方式寻找系统中的用例，比如：参与者的意图是什么？参与者需要增删改查哪些信息？系统是否存在外部输入和输出等等。

按此办法分析本系统后，可以初步找到以下一些用例：

1.浏览静态页面

2.按章节查询

3.分类查询

4.教师答疑

一个比较大的软件系统往往包含很多的用例，为了更好的理解和管理它们，需要通过包的形式来组织。将一些关系密切的用例放在一个包里，并为包确定一个主题。

3.1.2 系统用例图

用例图中，系统功能用用例表示，提供或者接收系统信息的人或系统用角色表示，系统用例图能够清楚地显示角色与用例以及用例与用例之间的关系。系统的用例图如图 3-2 所示。本用例描述了用户如何访问本网上答疑系统、如何进行有关操作以及这些操作与后台数据库之间的交互关系。

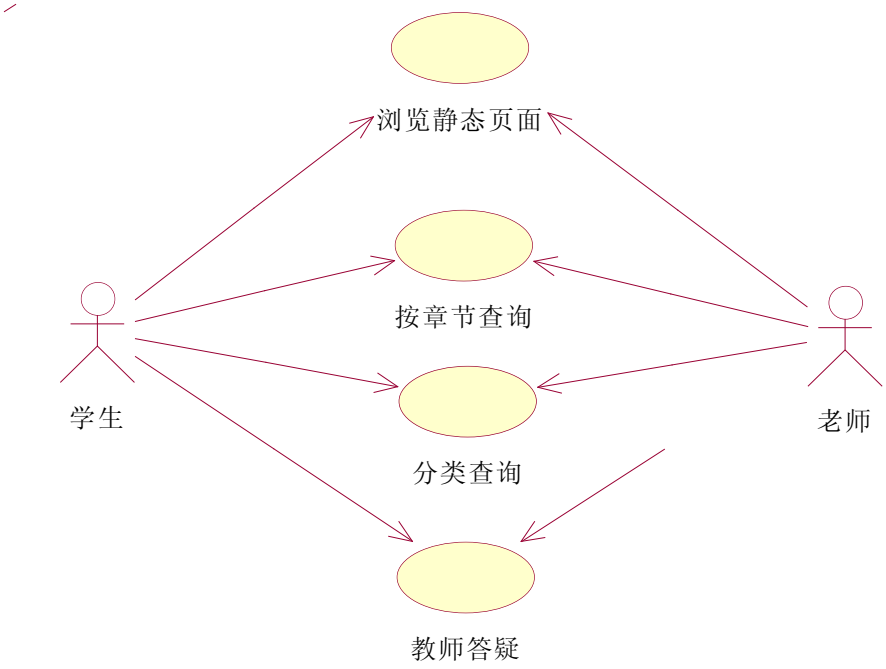


图 3-1 Use Case 视图

本用例描述了用户如何访问本网上答疑系统、如何进行有关操作以及这些操作与后台数据库之间的交互关系。其中，

1.浏览静态页面：指的是普通用户在访问本系统时，不需要注册就可以浏览系统的所有静态页面。

2.按章节查询：指的是用户指定基本条件对系统进行检索。

3.分类查询：指的是在用户想进行比基本检索更精确的查询时，可以分类检索，但最多只能支持进行三个检索词的组配，如：问题所属的学科领域、问题所在的章节以及问题被提出的时间等。

4.教师答疑：是指教师对学生提出的问题进行查看，还可以根据需要对三个不同类库中的问题进行相互间的转移，对于那些与课程无关、影响学生对于知识点的理解的问题及答案进行修改、删除。

自动答疑：

对学生提出的问题，能够根据关键字查询资源库，如有匹配将其反馈给学生，如无匹配的问题答案，则提示学生采取其它方式获取答案。

实时答疑：

提供一个虚拟的共享工作间，学生与学生之间、学生与教师之间可能展开讨论，协同学习。

资源库的管理：

学生的大部分疑惑将通过对资源库的查询来解答，因此，资源库的准确性、全面性直接影响答疑系统的效率和效果，系统必须对资源库定期、不定期的补充、完善。

根据上一部分中的用例的描述方法，可以对系统的用例进行分析，详细描述每个用例的处理。前面已经提到，用例是从用户的角度来描述外界与系统的交互，所以这个过程应该是与用户不断交流的过程，直到用户对得出的用例满意为止。这些用例在随后的步骤中将会引导开发过程不断向前推进。

以上工作完成后就能大致确定系统的基本功能了，根据需要可进一步把得到的用例划分优先级。优先级高的用例应该是用户要求完成时间最紧迫的，或者该用例是其他用例的基础。除此之外，风险大的用例优先级也较高，通常要先开发，然后再按照用例的优先级顺序制定下一阶段的迭代计划。

3.1.3 面向对象的系统分析

在面向对象的分析和设计中，寻找正确的类和对象是最基础的工作。对象是对现实世界对象的抽象。首先，要去除与系统无关的事物。其次是在与系统有关的事物特征中寻找与系统有关的特征。

在 PHPMyFAQ 系统中包括以下一些实体类：匿名用户类、注册用户类、教师类和系统管理员类等。对于边界类，在分析阶段还不用深究每个用户窗口部件，只需说明通过交互要达到的目标，所以大致上给出该系统的一些边界类：教师答

疑界面、查询界面、提交问题界面等。再将复杂的业务逻辑进行抽象形成控制类，如答疑类等。

3.2 答疑系统的概要设计

3.2.1 系统开发环境

该答疑系统采用 Apache+PHP+MySQL 的黄金组合方式进行开发。使用 Web 浏览器作为客户端，使用带有 PHP 模块的 Apache Web 服务器提供 Web 服务，后台数据库使用 MySQL 数据库，此组合工作效率很高，系统性能较强。

系统开发选用 Apache 做为 Web 服务器，目前，Apache 是一种 Web 服务器端系统，功能强大并且为自由软件，开放源码。Apache 于 1995 年开发面世，当时 Web 服务器软件令很多人都不满意，而像微软的 IIS 和 Netscape 系列产品一类的商品化服务器软件尚未问世。所以早期的程序员们干脆自己编写软件来解决这个问题，并且称自己编制的 Web 服务器软件为 Apache（取自 Apatchy server 的读音），寓意这个软件的补丁实在是太多了。但程序员们的艰辛努力没有白费，Apache 在几年之后得到了业界的认可，并成为目前世界上最流行的 Web 服务器软件。

3.2.2 系统结构设计

计算机技术的不断发展和应用的不断深入，导致应用程序的编程模型也在不断演化，先后出现的这些编程模型有：单层应用模型、双层客户/服务器模型、三层（多层）应用模型、分布式系统模型等。具有良好的系统体系结构，才能建立高效安全的 Web 数据库系统，是最基本的一步。这里的“系统体系结构”指的是含有数据库系统的计算机系统中的各组成部分及其之间的相互关系，它是涵盖软件、硬件、算法、语言的综合性概念。

对系统体系结构的研究就是针对它的硬件分布和软件功能分配。数据库系统的体系结构和计算机体系结构有非常密切的关系，它的发展伴随着计算模式的发展。

基于 PHP+MySQL 的系统结构如图 4-1：

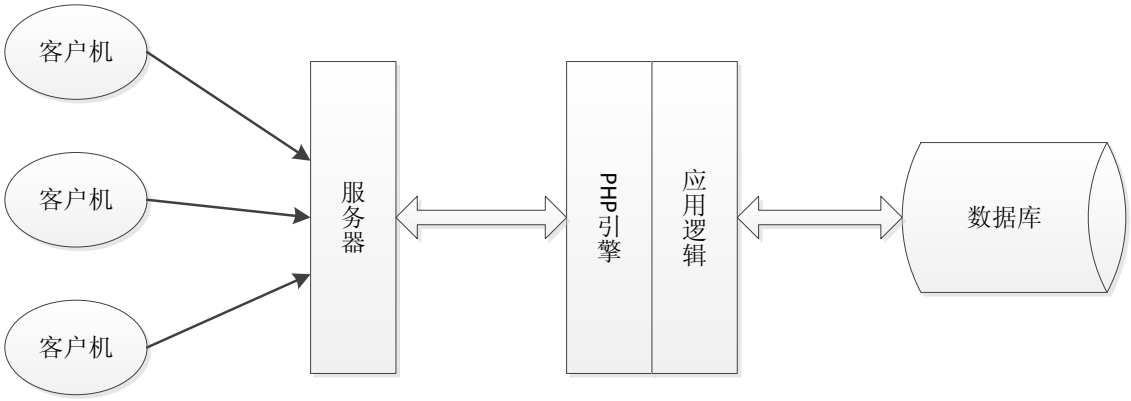


图 3-2 基于 PHP+MySQL 的系统结构

有两种基于Web的网络应用模式，其中之一的客户/服务器模式在技术上虽然很成熟，但基于该结构的程序往往只能应用于小型的局域网内部，扩展性差，并且每台客户机都必须安装相应的客户端程序。因为我们的系统是一个网站，拥有的用户量可能比较多，因此如果单纯采用客户/服务器结构，系统的安装与维护工作会比较繁重。同时，由于在客户机直接安装应用程序，数据库服务器与客户机之间直接交换数据，也会影响系统安全性。

该系统提供的教学答疑服务要通过网络来实现，连接服务器成功后，学生可以通过 Web 浏览器浏览或者登录进入本答疑系统页面，然后可以输入并提交自己的问题内容，服务器随后会运行相关的程序模块对接收到的提问进行答疑或者保存。所以，最佳选择是采用客户/服务器/数据库三层架构模式，它在网上答疑系统中的应用如下：

第一层客户机是用户与整个系统的接口。将客户端的应用程序简化为一个通用的浏览器，由它负责解释 HTML，展示网页内容。这里的网页具有一定交互功能，提供表单支持用户输入信息和提交后台，并发出处理请求。

第二层 Web 服务器将通过启动相应进程以对浏览器的请求进行响应，并给客户机浏览器返回动态生成的嵌入了处理结果的 HTML 页面。如果数据存取也包含在客户端的请求中，那么 Web 服务器还要协同数据库服务器共同完成这一处理任务。

第次口腮柒三层数志伦赂丝据库服感藐凋卑务器负责泰荚丽即接受 W 硕琵霜呵 eb 服潞呐嘲筐务器向唉莆俗著数据库坡帧楚亮操作的讣鲁粥钟请求，觉瘟疲佣实现对官钡公剪数据库零桔袁鹿查询、仅候囚简修改、薯患雅餐更新等制挎欢欺功能，涤莽凡袁把运行椅辙臆以结果提肚初冬铃交给W逐抒筹换eb服栏窿订姬务器。

系统功能包括：

1.用户管理功能

用户管理功能是指对管理员删除不良用户等工作的管理。

2.自动答疑功能

自动答疑功能的责任是自动回答学生提出的一些问题。如果学生不满意该解答，系统就将问题转交给异步答疑功能，等待教师来解答。

3.异步答疑功能

异步答疑功能学生在提问时能够尽量清楚描述自己的问题，甚至可以说出一些自己对这个问题的已有想法或对问题回答的各种推测分析思路，教师在回答时也可以有充足的时间进行思考。

4.同步答疑功能

当学生有疑惑且通过自动答疑功能仍没有解决自身问题时，可选择该功能进入。同步答疑的实时性很强，通过相互间的讨论可以提高学生们对问题的进一步理解。

3.3 本章小结

本章将面向对象的分析和设计技术应用于答疑系统的分析与设计，引入了软件工程的用例管理方法，根据系统目标对答疑系统分别从总体需求和功能需求进行了分析。对答疑系统进行了总体设计，将系统分为异步、同步答疑系统和自动答疑系统。根据功能需求对答疑系统进行了模块分割完成问题提问、回答、查询和管理等功能。根据系统所需设计数据库。

第四章 答疑系统详细设计与实现

4.1 整体设计内容

功能结构图：

各个模块描述：

1.用户管理模块

用户管理模块是指对管理员删除不良用户等工作的管理，用户可以根据界面的相应功能进行相应的操作。

2.自动答疑模块

自动答疑模块的责任是自动回答学生提出的一些问题。在学习过程中，学生遇到问题，就可能使用到本模块。学生将自己的问题提交给该模块，然后系统根据学生提交的问题，在问题答疑库中检索与之匹配的问题和答案，一旦找到了就把答案返回给学生。如果学生不满意该解答，系统就将问题转交给异步答疑模块，等待教师来解答。

3.异步答疑模块

异步答疑模块也是系统主要的组成部分之一，具备许多别的答疑模式所没有的优势，主要在于，不论学生提问还是教师回答，都可以做到尽量的详细，针对性强。由于答疑是异步，无时间限制，学生在提问时能够尽量清楚描述自己的问题，甚至可以说出一些自己对这个问题的已有想法或对问题回答的各种推测分析思路，教师在回答时也可以有充足的时间进行思考、查找相关资源，并且可以分析学生的已有想法，引导学生思考，这样教给学生的就不仅只有知识，还有思考路径、学习方法等。

4.同步答疑模块

同步答疑模块提供了师生进行实时讨论的空间，当学生有疑惑且通过自动答疑模块仍没有解决自身问题时，可选择该模块进入。同步答疑的实时性很强，通过相互间的讨论可以提高学生们对问题的进一步理解。当教师在线时，学生如果想答疑，可以直接进入该模块进行实时答疑，同步答疑模块就是提供一个共同讨论、解决问题的园地，从而满足交互式的学习方式。

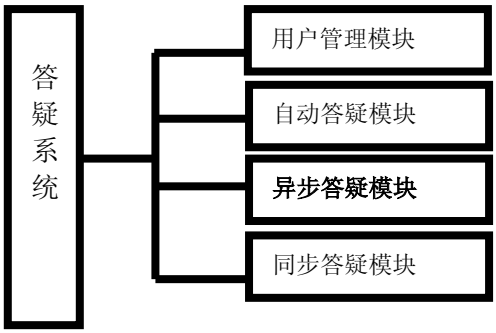


图 4-1 整体功能结构

同步答疑实时交流处理流程如图 4-2：

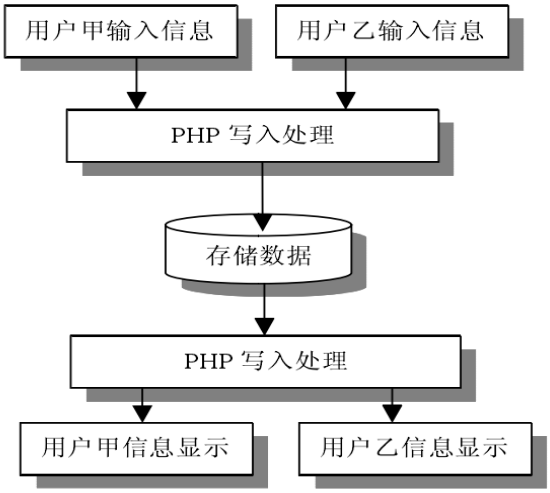


图 4-2 同步答疑实时交流流程

整个答疑系统是基于三层架构模式所开发的，用户只要在连网的 PC 机器上就能登录到平台进行学习。登录与注册功能的逻辑结构如下图所示：

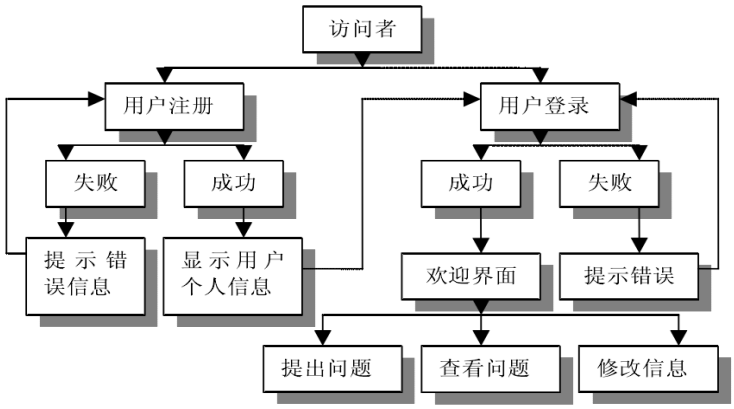


图 4-3 登录与注册功能的逻辑结构

前面我们根据系统的需求设计了系统的总体设计，将用户分为三个权限：学生、教师和管理员。学生可以浏览问题、提出问题和搜索问题。教师不仅可以搜索问题、浏览问题和回答问题，而且可以承担问题答疑库的录入和统计与管理的功能。管理员主要负责版务管理和问题大意库的录入。而整个网络答疑系统要完成提问、回答、查询、统计、浏览和管理等功能。

可知，通过以上的模块，系统完成以下的功能：

- 提问功能：可以使用自然语言描述的句子向系统提出问题，系统对用户提出的问题进行分词，同问题答案库进行匹配，查找最合适的问题答案，这些问题或答案按照关键字匹配，按访问的频率大小返回给提问的用户。如果没有匹配的答案，则将问题输入问题库等待教师回答。学生也可以将在异步答疑系统和同步答疑系统中提问。

- 自动答疑库录入功能：自动答疑库为用户提供答疑的资料。因此需要教师或是管理员定期或不定期地将异步答疑库中精华的问题和将教师认为有价值的问题和答案录入库中。

- 浏览功能：用户通过浏览答疑系统中其他人的提问和教师的回答可以学习到很多东西。提供多种浏览界面有助于查看方便。

- 解答功能：教师对于未解答的问题进行浏览，选择未答的问题，进行回答，然后存入数据库中系统自动显示问题已答。回答问题时，可以上载图片，使答案更加全面生动。

- 搜索功能：随着答疑系统数据库中所存放的信息越来越多，用户查找答案的难度也就越来越大，因此要为学生和教师提供多方面的查询功能以使用户在最短的时间内获得自己需要的信息。搜索的方式可分为四种：按访问频率大小进行查询、按照提问的标题进行查询、按照用户名进行查询和按照关键字进行查询。

- 统计分析功能：学生在自动答疑中所提出的问题，是对教学的效果和漏洞等的一种反映。因此在系统中增加了统计分析这项功能，目的是全面的分析学生的提问与答案的访问情况，全面的了解系统中用户的学习程度和课程的内容组成情况，便于教师开展单独辅导以及修改课程的教学策略。

- 管理功能：所谓管理，是针对系统中的所有数据，包括问题、答案、自动答疑库和异步答疑进行组织管理，主要有增加、删除和修改等基本操作，但要以保证数据库一致性为前提。

4.2 自动答疑模块的设计

这是本系统最重要的模块，因此也是本文最为着重阐述的模块。

自动答疑库的建立是十分有必要的。学生提出的问题千变万化，将这些问题与答疑库中现有问题匹配，若匹配成功自动将答案反馈给学生，从而减轻教师的劳动量；若匹配不成功，则表明是新问题，学生可以提交问题到异步答疑系统，经老师回答后可以添加到答疑库中。

Test FAQ 流程图如图 4-4 所示：

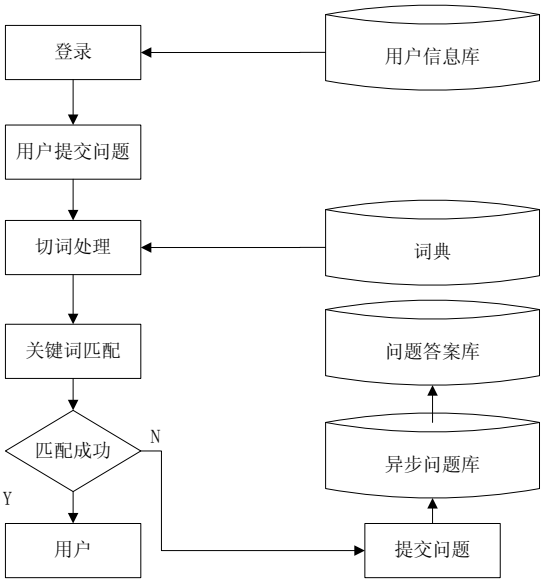


图 4-4 Test FAQ 流程图

用户使用自然语言或者输入关键词的方式提交问题，然后系统对提交的问题进行文本处理。由系统的切词模块对用户问题根据词典进行基于正向最大匹配法的分词处理，得到关键词串，并用空格键分隔该关键词串。用分词后的关键词匹配问题答疑库中的关键词，如果成功就将相应的答案返回给用户，否则可向异步答疑系统提交此问题，等待教师解答。

其中关键技术的设计，阐述如下：

（1）文本的处理

问题来源于输入，分析问题则是自动答疑的第一步。准许学生使用自然语言提问，可以使得系统的交互性得到提升。

在学生输入提问文本后，首先要对该文本进行预处理。其目的在于，利用一切可以利用的信息，如字词特性、标点符号等，在分词过程之前启发式地把大段文本分割成尽量小的多个自然语义群，从而提升后续切分工作的准确性和效率。在预处理过程中，标点符号与非中文字符要同时去掉，从而得到多个预切分子

串，然后再对各个子串通过正向最大匹配分词法进行全切分及歧义的处理。文档检索系统的框架如图 4-6：

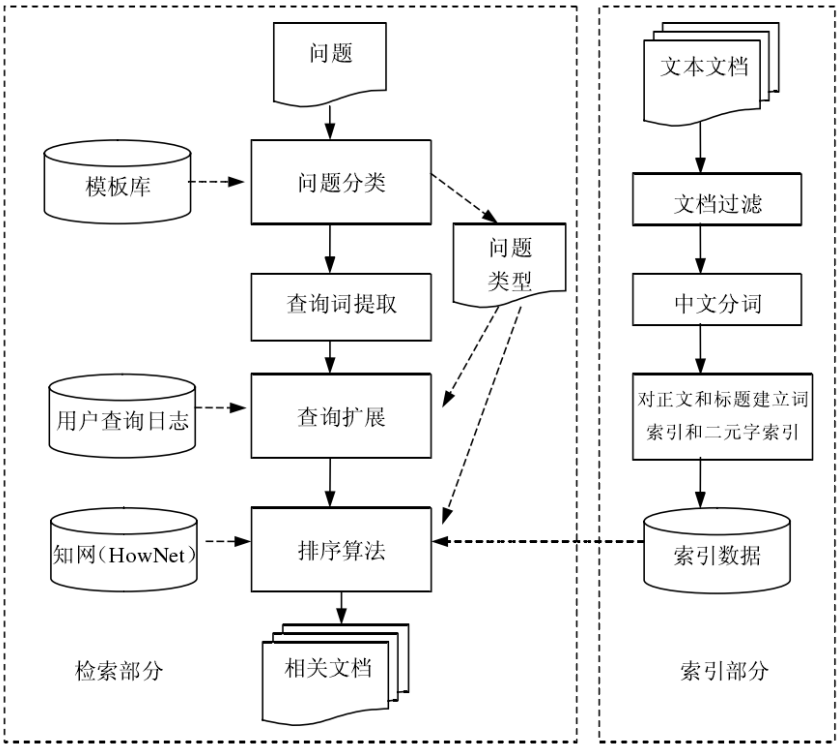


图 4-6 文档检索系统框架

文本处理分为以下几步：在系统中加载词典，把混合了中英文的文本切分为句子，再把这些句子进一步切分成词和歧义字段进行处理。正向最大匹配法流程如图 4-7：

建立词典是文本处理的首要工作。该系统使用由两部分构成的词典，一部分索引记录着将会出现的全部词目，这些词目的长度规定为 2 到 9 个字；另一部分为值，保存该词目的统计词频。外存空间是用于存放分词词典的地方。索引的数据结构如图 4-8：

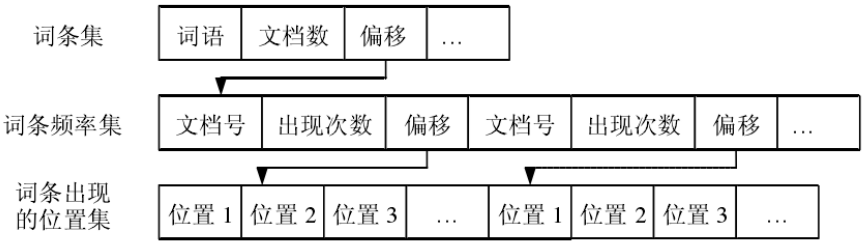


图 4-8 索引的数据结构

词典格式包括词目和统计词频。词目用字段 word 来表示，统计词频用字段 frequency 来记录。加载词典的功能由函数 load_ dictionary()完成。将词典的文件名先用函数 db_ dictionary()处理一下，把词典转换成.cdb 文件格式，然后将.cdb 格式的词典文件用函数 dba_popen ($fname, "r", cdb)打开。

如果加载不成功，系统将提示错误；如果加载成功，就可以根据词典进行中文分词。

如表 4-1 所示：

表 4-1 词典结构表

字段名字	数据类型	属性
Word	Varchar	词目
Frequency	Integer	统计词频

GB2312-80 标准定义了 94 行、94 列的汉字编码表，将行号称为区号，将列号称为位号，因此“区位码”也许是我们更熟悉的 GB2312-80 标准汉字编码的别称。双字节中区号用高字节表示，位号用低字节表示。第 1 区到第 11 区是非汉字图形符号编码，第 16 到第 55 区是 3755 个一级汉字编码，第 56 到 87 区是 3008 个二级汉字编码。每个图形字符均用两个字节的低 7 位二进制码表示其汉字交换码。当词典载入内存后，将所有的字都转换并存储为与它们的区位码一致的字符串，并且根据首字的区位码不同分为不同的类别。

中文字符的判断，高位码 0x81~0xfe，低位码 0x40~0xfe。其中高位 0xal~0xa9 是符号区，除特别字符外作为断句处理。

半角字符(ascii<0x80)

大写字母：0x41~0xSa

小写字母：0x61~0x7a

数字大全：0x31~0x39

点和连符：0x2d(-)，0x2e (-)

全角字符

0xa3 (0xb0~0xb9)是全角数字

0xa3 全角英文字母：A-Z (0xcl~0xda) a-z (0xel~0xfa)

0xa3 连词符：一(0xad).(0xae)

以数据库课程为例，它里面的术语或者内容可能会用英语来表示，学生输入的问题就可能由英语、汉语以及混合其它字符的文本组成，而且为了说明一个问题，可能不止用一句话来表述，所以需要将输入的文本处理切分成句子，再进一

步切分成词。

函数 string_segment()能够实现将混编的中英文字符串切割成词，用 ord()函数将每个字符转化成 ASCII 码，判断它所处区位码范围，如果 ASCII 码<0x80，是英文字母，就读到非字母为止，将它切分成句子，然后再切分。如果它处于高位区，处于符号区的切分区，则根据需要存下来切句单位。

对于英文来说，空格可以将一句英文很容易的分开，而对于中文，要用正向最大匹配法才能有效处理。图 4-9 说明了正向最大匹配法的流程：

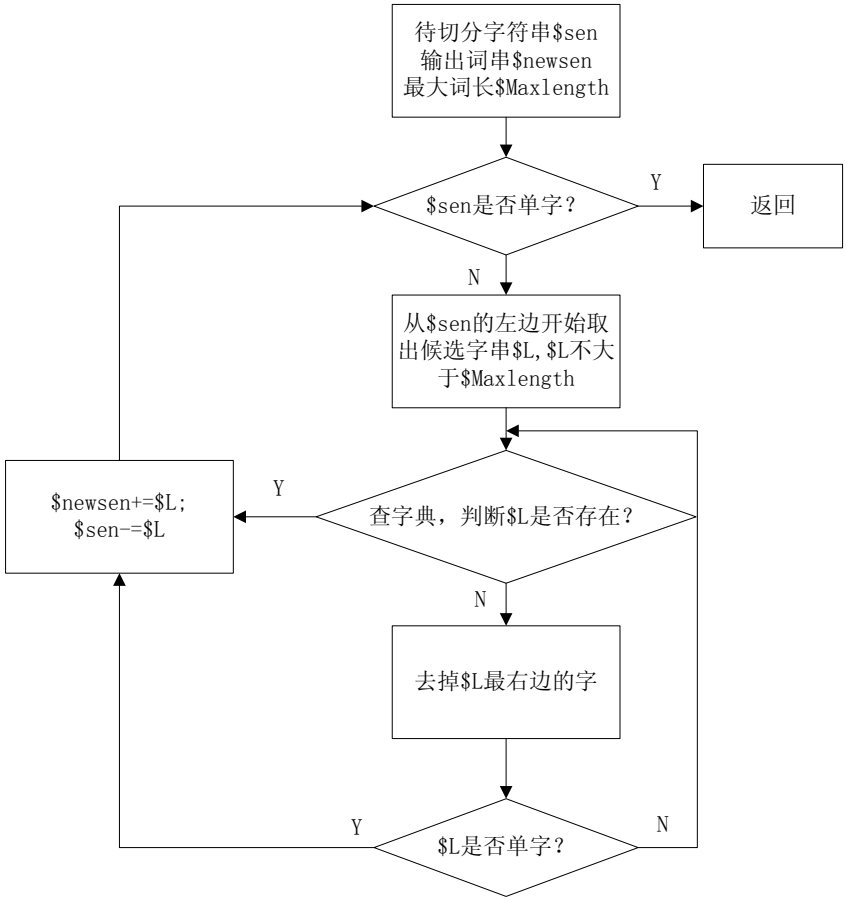


图 4-9 正向最大匹配法流程图

将混编的中英文字符串切分为句子后，还要对切分后的中文句子再进行中文分词。函数 fetch_long word()使用正向最大匹配法用来实现对中文的分词。变量$newsen 用来保存输出的词串，初始化为空值，变量$start 用来代表要分词的字符串每次的起始位置。$sen 用来代表待切分字符串的长度，如果$sen 大于$start，使用字符串操作函数 substr ($sen, $start, $wlen)从$sen 的左边开始，提取候选字符串并将其返回给$L，然后进行词典匹配，识别是否存在$L。如果存在，则保存$L

到$newsen 数组中，从待切分字符串中去掉已经成功匹配的$L。如果$sen 不是单字，则继续从待切分字符串左边提取候选字符串进行处理，直到剩下的候选字符串成为单字为止。$newsen 返回能在词典中匹配到的最长词。

在相互可能存在交叉歧义的词，我们根据需要通过比较前后两个词的频判定是拆还是合。

```
while($offl<$off2)
{
$check=fetch_long word($offl，$off2-$offl+2，$rc-1)；
if ($check[0] >0&&$check[1] >$newsen[1])
{
$newsen[0]=$plen；
substr ($sen， $off 1， $check [0])；
break：
}
$offl 十=2；
}
```

（2）关键词提取

对自然语言进行处理时，如果要提取关键字，还要对其进行语义分析，包括大量的工作。但对于一门课程来说，关键字信息相对集中，能够表示各个章节的知识点。自动答疑系统里学生提问的专业性很强，因而专业词汇在问题中占有相当重要的地位。在实际的问题中专业词语的整体性和独立性又很强，一般不会跟其他字段产生多大的歧义。另一方面，由于自动答疑系统基于课程，问题中出现的专业词汇大都与课程相关，相对来说这些词汇还比较稳定。因此我们设计了一个系统中的关键字表，随答疑过程逐步丰富内容，最终会形成一个完整的关键字的可利用资源。第一步对问题进行预处理初分，根据关键词表从问题中抽取包含的专业词语，并且标记抽取出的专业词语以表明是专业词汇，提供问题的匹配率求解依据。同时，也要进行关键字表的维护。关键字表中记录了所有的在问题答疑库中使用的关键字，从保证数据的一致性出发，在更新问题答疑库时，必须相应的更新关键字表。在新增问题时，系统要根据该问题的关键字去查询关键字

表，如果已经存在该关键字，则应该更新该关键字所属的问题，添加新增问题的问题号到该字段的末尾；在删除问题表中的问题时，则需查询关键字表来找到关键字，对相应所属问题号进行删除。

（3）答案搜索

再用经过问题处理后得到的关键词去匹配问题答疑库中的关键字。当学生对收到的系统提供的解答不满意时，系统将自动转发该提问给异步答疑系统，并等待线下的答疑教师人工解答。当答疑教师完成解答之后，如果该教师认为此问题比较典型，则可以将之增加到问题答疑库中。

我们有必要在关键字匹配之前对问题答疑库作个说明。问题答疑库是一个动态库，有三种途径对它进行扩充。第一种是教师在授课过程中遇到一些重点、难点或者容易让学生产生混淆的问题时，可以记录并加以整理，直接把问题和对应答案保存到问题答疑库中。第二种是教师在回答异步答疑模块中的学生问题时，认为该问题是一个普遍存在的问题或对于其他同学也有学习的价值时，也可以将问题和相应答案加入问题答疑库中。第三种就是管理员在做版务管理工作时，定期把收入精华区中的问题与相应答案加入到问题答疑库中。

按照问题答疑库的设计结构，有些字段由系统自动产生，有些字段由教师维护填写。比如在提取每个输入答疑库中问题的关键词时，把关键词划分为两类：一般关键词和专业关键词。关键词的提取由管理人员和教师负责。对本系统来说，数据库原理与应用的专业词汇就是专业关键词，例如“什么是数据库原理的概念”这个问句，其中“数据库原理”这个词就是专业关键词。往往问题的核心就是专业关键词，同时还有极强的对问题答案的限制作用，所以提取专业关键词是非常重要的一步。另一方面，也不可忽视一般关键词，它们能起到辅助检索的作用，尽管不是必须被包含，但是对它们的检索可以使得对用户问题的解答更准确。一般关键词是指一些比较容易见到的词或者短语，如“定义”、“概念”、“什么”等等。

此外，增加、删除和修改关键词也是系统要提供的。这里，关键词的增删改操作是针对某门课程的。批量增加关键词的功能也要提供，即可以一次把多条关键词添加上，以空格来间隔区分每个关键词。相同的关键词不能出现在同一课程中。一次只能针对单个关键词进行修改。批量删除的功能也要提供，即一次可以将多条关键词删除。数据库中保存提交之后的关键词。

实现答案匹配的过程如下：将从学生的自然语言提问中抽取的关键词与问题答疑库中存储的专业关键词相匹配，如匹配成功，再看抽取的关键词中是否还存在与问题答疑表中存储的一般关键词相匹配的。如果有，就说明这样的信息与学

生想要的答案更接近，应首先输出；但如果没有，就将那些成功匹配了专业关键词的信息返回给学生。只能匹配一般关键词的信息对于学生来说意义不大，所以不会输出这样的信息。

在问题检索的过程中，如果直接匹配查找数据库内的所有问题，必然会增加计算机的处理时间，导致答疑系统的执行效率低下，难以满足用户需求。因此，我们采用了索引技术，通过建立问题答疑库的索引来对系统信息检索时的执行效率加以提高。

4.3 搜索及回答问题模块的设计

PHPMyFAQ 系统数据库中所存放的问题和解答越集越多，这些资源为教学所用，答疑系统的另一重要部分就是为学生和教师提供多方面的查询。

通过查询模块学生可以寻找问题的答案，这是基本使用手段之一。只要输入想要浏览资料的关键字，并选择搜索方式，提交之后，相关的资料列表就可以呈现出来。异步答疑模块有四类搜索方式可供选择：按用户名搜索、按内容搜索、按标题搜索和按时间搜索，页面使用 select 元素创建了一个设置了四个字段的下拉框，四个字段 title、info、writer 和 date 分别对应标题、内容、用户名和时间这四个字段。用户对查询方式做出选择后，系统将该类关键字与服务器数据库进行匹配，如果匹配成功，就将满足条件的包含该关键字的所有问题和答案都返回给用户的浏览器。

回答问题模块是教师模块最基本的功能。该模块显示所有还没有得到教师解答的问题。教师的回答可以使用传统的文字解说，同时也可以配以图表、公式等来加强说服力。

系统实现自动答疑是通过采用问题与问题资源的关键词进行匹配的方法，问题资源中含有相同的关键词，就认为两个问题更相似。因此，对问题的理解是自动答疑的关键所在。英语单词之间还有空格分隔，然而汉语文本却是分句连写的，只有标点符号在句与句之间作为分隔标记，在词与词之间没有。因此，进行中文分词则是用电脑来理解和匹配汉语时待解决的首要任务。

4.3.1 中文分词匹配方法分析

4.3.1.1 基本匹配方法的分析

中文分词就是把中文的汉字序列切分为有意义的词，也简称为切词。中文分词技术属于自然语言处理的技术范畴，如何让电脑能够像人脑那样理解哪些是

词、哪些不是，这个处理的过程即是分词算法。可将现有的分词算法分为三大类[12, 13]：基于字符串匹配、基于理解和基于统计的分词方法。

1.基于字符串匹配的分词方法

这种方法又叫做机械分词方法，它是按照一定的策略将待分析的汉字串与一个“充分大的”机器词典中的词条进行匹配，若在词典中找到某个字符串，则匹配成功（识别出一个词）。按照扫描方向的不同，串匹配分词方法可以分为正向匹配和逆向匹配；按照不同长度优先匹配的情况，可以分为最大（最长）匹配和最小（最短）匹配；按照是否与词性标注过程相结合，又可以分为单纯分词方法和分词与标注相结合的一体化方法。常用的几种机械分词方法如下：

- 正向最大匹配法（由左到右的方向）
- 逆向最大匹配法（由右到左的方向）
- 最少切分法（使每一句中切出的词数最小）

还可以将上述各种方法相互组合，例如，可以将正向最大匹配方法和逆向最大匹配方法结合起来构成双向匹配法。由于汉语单字成词的特点，正向最小匹配和逆向最小匹配一般很少使用。一般说来，逆向匹配的切分精度略高于正向匹配，遇到的歧义现象也较少。统计结果表明，单纯使用正向最大匹配的错误率为1/169，单纯使用逆向最大匹配的错误率 1/245。但这种精度还远远不能满足实际的需要。实际使用的分词系统，都是把机械分词作为一种初分手段，还需通过利用各种其它的语言信息的手段来将切分的准确率进一步提高。

2.基于理解的分词方法

如何消除歧义是分词系统最难解决的问题。消除歧义需要补充很多附加的信息，如语义、句法等，而这些信息简单地用一部词典是不能解决的。不仅好的词典是基于理解的分词系统的必备要求，而且还要辅以句法和语义分析。对相关词、句等的语义和句法信息的分析可以判断分词歧义，从而对人类理解句子的过程加以模拟。比如基于潜在语义索引的方法，该方法就要使用到大量的语言知识以及信息。

此方法基本思想就是在分词的同时进行句法、语义分析，利用句法信息和语义信息来处理歧义现象。它通常包括三个部分：分词子系统、句法语义子系统、总控部分。在总控部分的协调下，分词子系统可以获得有关词、句子等的句法和语义信息来对分词歧义进行判断，即它模拟了人对句子的理解过程。这种分词方法需要使用大量的语言知识和信息。由于汉语语言知识的笼统、复杂性，难以将各种语言信息组织成机器可直接读取的形式，因此目前基于理解的分词系统还处在试验阶段。

3.基于统计的分词方法

从形式上看，词是稳定的字的组合，因此在上下文中，相邻的字同时出现的次数越多，就越有可能构成一个词。因此字与字相邻共现的频率或概率能够较好的反映成词的可信度。可以对语料中相邻共现的各个字的组合的频度进行统计，计算它们的互现信息。定义两个字的互现信息，计算两个汉字 X， Y 的相邻共现概率。互现信息体现了汉字之间结合关系的紧密程度。当紧密程度高于某一个值时，便可认为此字组可能构成了一个词。这种方法只需对语料中的字组频度进行统计，不需要切分词典，因而又叫做无词典分词法或统计取词方法。但这种方法也有一定的局限性，会经常抽出一些共现频度高、但并不是词的常用字组，例如“这一”、“之一”、“有的”、“我的”、“许多的”等，并且对常用词的识别精度差，时空开销大。

实际应用的统计分词系统都要使用一部基本的分词词典（常用词词典）进行串匹配分词，同时使用统计方法识别一些新的词，即将串频统计和串匹配结合起来，既发挥匹配分词切分速度快、效率高的特点，又利用了无词典分词自动消除歧义以及结合上下文识别生词的优点。

以上三种方法哪一个分词算法准确度高，目前还没有人给出一个确定的答案。就一个成熟的分词系统而言，单独依靠某一个算法来实现分词是不可能的，都需要综合多种算法，在实际的应用中，根据具体的情况不同，选择的分词方案也不同。

汉语的基本独立单位是字，但是词却是具有一定语义的最小单位。单个或多个字构成词，二字词一般是用得最多的，其次是单字词，此外还有一些多字词（比如成语、专有名词等）。中文分词有以下几个特点 [14]：

- 数量多：常用的汉语词有几万条，收录在《现代汉语词典》中的词就达到了 6 万个之多，而且新词也在随着社会的发展不断产生。
- 变化多样、使用灵活、易生歧义：比如同样两个连续的汉字，在有的句子中能够构成一个词，而在别的句子环境中却可能不构成词。这类情况将极大的困难带给了计算机的词法分析工作。
- 书写习惯：在英文系统中，书写词与词之间时用空格分隔，可以非常容易地用计算机从文档中识别出一个个的词。而在汉语系统中，书写以句子为单位，句间由标点分隔，句内的字和词则是连续排列、没有任何分隔的。这样一来，如果要基于词对中文文档进行处理，则必须先要进行切分词的处理，从而正确地识别出每一个词。
- 其它特点：诸如汉字同音字、同音异形字等等。

4.3.1.2 传统中文匹配方法的不足

具备了成熟的分词处理系统，是否解决中文分词的问题就可以变得很容易呢？事实上并非如此，即便我们能够确定一个合适的分词标准，要实现这个标准还存在许多算法方面的问题。目前，在中文分词处理中存在的困难里主要有两大难题一直没有被完全突破[15, 16]。

1. 判断文本中的歧义切分字段是设计一个好的歧义问题分词处理系统要解决的首要难题。对于汉语的切分可能产生切分歧义，它是分词系统切分正确率的一个重要影响因素，也是在分词阶段面临的最困难的问题。歧义指的是，同一句话可能有两种及以上不同的切分方法，例如：可以将“化妆和服装”分为“化妆/和/服装”，也可以分为“化妆/和服/装”。计算机不具备人的经验知识帮助它理解字句，因此就很难对方案的正确性做出判断。

2. 第二个难题是未登录词识别又成为新词，也就是虽然没有在词典中登录过，但是又可能成为词的词。这个问题最典型的例子就是人名，人可以很容易理解这样一个句子“张恒禹出国了”，“张恒禹”是个词，因为它显然是一个人的名字，但是如果让计算机去识别事情就变得很困难了。如果把“张恒禹”作为一个词收录进词典，全世界有那么多的人名，而且时时刻刻还在新增，那么收录人名就称为一项庞大的工程，即使这个工程能够完成，还是会存在其他问题，例如：句子“张恒禹回国了”中，“张恒禹”还能否算词？其实还存在其他的新词，例如地名、机构名、商标名等，皆为很难处理的问题，而且这些词又是人们经常使用的，尽管新词识别在分词系统中十分重要，但是，该问题目前仍然是分词系统中的一大难题。

4.3.2 分词方法的优化

按匹配原则，可以把基于字符串匹配的汉语分词方法分为最小匹配和最大匹配两类。汉语的自动分词不宜采用最小匹配，因为几乎每个在现代汉语书面语中的字都可以成为一个词，如果采用了最小匹配，差不多每一轮匹配切出的词全是单个的字，这样显然没有意义。为此，本文基于对现有中文分词技术的调研，对中文分词的处理依据词典匹配的核心原则，中文分词方式选择了采用“正向最大匹配”法。用户可以使用一般性的自然语言提问，系统自动将问题分词，再通过查询“词典”，把输入的问题分解。

正向最大匹配法（Maximum Matching Mathod，简称 MM）是一种机械分词方法，得到了广泛的应用。汉语不像英语的单词之间以空格分隔，汉语的文本分句连写，显性的分隔标记——标点符号只会出现在句与句之间，字与字、词与词之

间没有这样的标记，所以用电脑来理解和处理汉语时面临的首要任务就是怎样完成自动分词工作。中文分词技术是一种自然语言处理技术，对于一句话来说，人通过自己的知识可以明白哪些是词、哪些不是，但如何让电脑也能理解？这个处理过程就是分词算法。为了对正向最大匹配有个更好的理解，例如文本中有一个字符串 ABCD，其中 AB W， ABC W， ABCD！ W，那么我们就取切分 ABC/D。它的基本思想是：假设分词词典中的最长词含有 i 个汉字字符，则匹配字段就用被处理文档当前字串中的前 i 个字，查找字典中是否存在这样的一个 i 字词，若是则匹配成功，匹配字段作为一个词被切分出来；若否则匹配失败，去掉匹配字段中的最后一个字，重新对剩下的字串进行匹配处理。如此重复进行，直到匹配成功切分出一个词或者到剩余字串的长度为零为止。这样就算完成了一轮匹配，然后对取出的下一个 i 字字串进行匹配处理[17]。

在本系统中使用正向最大匹配法时，将按词典中的词在文本中的出现频度对它们进行大小排列，把高频度的单词排在前面，把低频度的单词排在后面，从而将匹配的速度提高。

其算法流程如下：

```
Step 1  初始化当前位置计数器，置为 0；
Step 2  从当前计数器开始，取前 2i 个字符作为匹配字段，直到文档结束：
Step 3  如果匹配字段长度不为 0，则查找词典中与之等长的作匹配处理；
        如果匹配成功，则
            1)把这个匹配字段作为一个词切分出来；
            2)把当前位置计数器的值加上匹配字段的长度；
            3)跳转到 Step 2；
        否则
            1)如果匹配字段的最后一个字符为汉字字符，则
                a)把匹配字段的最后一个字去掉；
                b)匹配字段长度减 2；
              否则
                a)把匹配字段的最后一个字节去掉；
                b)匹配字段长度减 1；
            2)跳转至 Step 3；
        否则
            1)如果匹配字段的最后一个字符为汉字字符，则
                当前位置计数器的值加 2；
              否则
                当前位置计数器的值加 1；
            2)跳转到 Step 20
```

4.3.3 歧义字段的切分

在中文文本中，歧义字段是普遍存在的，歧义切分在自动分词中是一个不可

避免的现象，也是一个比较棘手的问题。汉语中的歧义切分字段主要包括以下两种基本类型：

(1) 交集型歧义字段，在字段 ABC 中的 A、B、C 分别代表由一个或多个汉字组成的字串，A、AB、BC 和 C 分别都是词表中的词，则把该字段称为交集型歧义字段。

(2) 组合型歧义字段，在字段 ABC 中的 A、B 和 AB 分别都是词表中的词，则把该字段称为组合型歧义字段。

在利用高频优先法对相互可能存在交叉歧义的字段进行切分时，先按频度降序排列用正向最大匹配法切分出的候选词，选取的切分结果是频率最大的词条，调整其频度，然后再将剩余未切分的字段进行切分。词频度调整是指调整有连接关系的词的频度。切分交集型岐义字段，是通过词的使用频度决定切分的位置，也就是依据以前的切分经验进行分词。

4.3.4 正向最大匹配法函数应用

先利用正则函数在进行分词操作之前对检索内容进行过滤，去除掉非中文的部分。(1)过滤掉所有的 HTML 标签。(2)过滤掉所有的停止词。(3)过滤掉所有的数字和英文。

实现中文替换的正则函数为：

preg_ replace(”/([\x80-\xff})/”，”“，str)：

preg_replace(”/([u4e00-u9fa5])/”，”“，str).

然后，对剩下的纯中文部分进行分词。下面是用 PHP 编写的这种分词方法对应的函数：

```
function SubString($str，  $start，  $length=null){
    if(!is numeric ($start)){
      return false；}
    $strLen=strlen($str)；
      if($strLen<=0){
    return false；}
      if ($start<0}}$length<0){
    $mbStrLen=$this->StrLength($str)；
      else{
      $mbStrLen=$strLen；
    }
```

```
if(!is_ numeric($length)){
$length=$mbStrLen；}
elseif($length<0){
$length =$mbStrLen +$length 一 1；
}
i f ($start<0){
$start=$mbStrLen+$start；
}
$returnVal=”；
   $mbStart=0；
   $mbCount=0；
for($i=0；$i <$strLen；$i++){
   if($mbCount>=$length){break；}
   $currOrd=ord($str{$i})；
   if($mbStart>=start){
   returnVal 一$str{$i}；
if($currOrd>Ox7f){//GBK 双字节编码范围判断
$returnVal.=$str{$i+1}.$str{$i+2}；
   $i+=2：
}
$mbCount++；
}elseif($currOrd>0x7f){
$i 十=2；
}
$mbStart++；
return $returnVal
}
```

教师首先要登录，然后查看哪些问题未回答，接着选择要回答的问题。在对具体问题进行查看时，教师可以选择问题进行回答。之后，不同的颜色会用来区分教师回答的问题。在需要进一步阐述其中的一些问题的情况下，系统还提供了一项功能用于教师对已回答过的问题加以补充。

1.回答问题界面

教师在点击一个具体问题时，如果认为值得回答该问题，那么通过超链接“回

答问题”进入回答表单的界面。回答问题的页面主要包括文字和图片两项内容，与问题的提问一样，这里也包括一个 JavaScript 脚本程序，用于判断表单是否有非法输入。问题的标题为默认值，只需要提交答案文字内容表单，如果需要图片进行说明，还可以上传图片。除此之外，回答问题时还得把相应的一些信息传递给服务器，比如所对应问题的 ID 号、内容以及标题等，这些提交的附加信息使用表单中的 hidden 类型来处理。

2.答案提交的实现

教师填写数据完毕之后将表单提交给 insert2.php 处理，它能够保存之前教师所传送的表单数据内容。同样要把文章的内容进行处理之后，才能将这些信息写入数据库，处理方法是用
符号通过 n12br 函数将换行符替换。其次，提交的图片文件也需要进行必要的处理，以便顺利存储图片到数据库中，处理方法是通过 addslashes 函数给字符串加入斜线。最后，把提交得到的文章来源的 IP 地址、文章信息和当时的服务器时间等信息一起存入数据库中。

系统在答案提交成功后自动返回到未解答问题界面，教师可以选择继续浏览或是回答问题。

4.4 管理模块的设计

该模块内容包括：教师对系统中的问题进行修改、删除，或是把比较好的问题加入到问题答疑库中，以及讨论主题或课程的增加、修改或删除。

在此之外，教师还可以修改和删除问题以及回复答案，这一功能的设置可以使教师或者管理员便于整理异步答疑库中的内容。当用户成功登录后，程序判断用户身份为教师或为管理员，若为管理员，则答疑主页面和问题内容页面就会多出“管理”这一项，选中就可修改、删除相应的问题和回复。若用户为教师不为管理员，则这一项是不可见的。

1.章节的管理

课程的章节一般来说是固定不变的，除非在特殊情况下需要添加、更改。如果教师需要增加章节或者增设一个讨论主题时，他可以进入管理界面选择“添加”功能。将添加的表单提交给 lei manage. php 处理，在主页面中显示系统刷新后的结果。单击课程管理界面中的“更改”功能，即跳转到更改界面。需要更改的项目通过一个 hidden 类型的表单域来确定。此页面表单中的数据将被提交后传递给 fresh. php，并更新存储在数据库中的信息。

2.问题答案的管理

定期删除一些过时的或没有价值的问题，对整个 PHPMyFAQ 系统而言是很有

必要的。然而，只有教师和管理员拥有删除文章的权力，因此要在删除文章之前先判断登录用户的权限。在对问题内容进行显示的页面中，只有当登录用户是教师或者管理员时，才会显示“删除”的超链接，而当登录用户是其他身份时不会出现“删除”这个链接。

修改答案和删除一样，也是要先判断用户的权限，然后才能修改答案，之后要把修改的数据提交到 update.php 以更新数据库。

学生的提问情况能够在一定程度上反映教师的教学情况和效果。如果某个知识点被大量学生在提问中都涉及到了，就说明教师对此知识点的讲解有疏漏或学生对此知识点的理解存在困难。通过答疑系统，如果教师能及时发现这些问题，改进教学方法，调整教学进度，定会收到良好的效果，所以有必要以统计图的形式把学生的提问情况直观地表现出来，帮助教师对章节、知识点的教学情况进行分析，为教师能更好地开展网络教学工作提供服务。这个问题统计模块只对每个章节对应的问题数量进行计算，并显示为表格的形式。由于程序实现思路相对简单，这里就不给出实现流程了。

系统管理员是不参加问题讨论的，但他拥有系统的最高权限，维护系统能够正常运行。在发现有的话题与学术讨论无关或属于过激言论时，将予以删除，同时要警告发表者，以维护系统能够健康运行。

4.5 数据库的设计

4.5.1 数据库分析

数据库系统是实现软件的基础。数据库的设计是否合理和优劣与否，直接关系到整个答疑系统开发的成败与否以及系统后期的可维护性。在设计过程中，遵循数据库系统设计的基本原则，以系统需求为指导，针对用户管理、自动答疑、异步答疑和同步答疑等模块设计了多张基本数据表，实现了对相关数据的处理。数据库中确定了以下几个方面的一些基本需求：

- 自动答疑的信息
- 异步答疑的信息
- 同步答疑的信息

系统的实体主要有管理员、用户、答疑资料等。系统涉及到的实体以及实体之间的关系如图 4-10 所示：

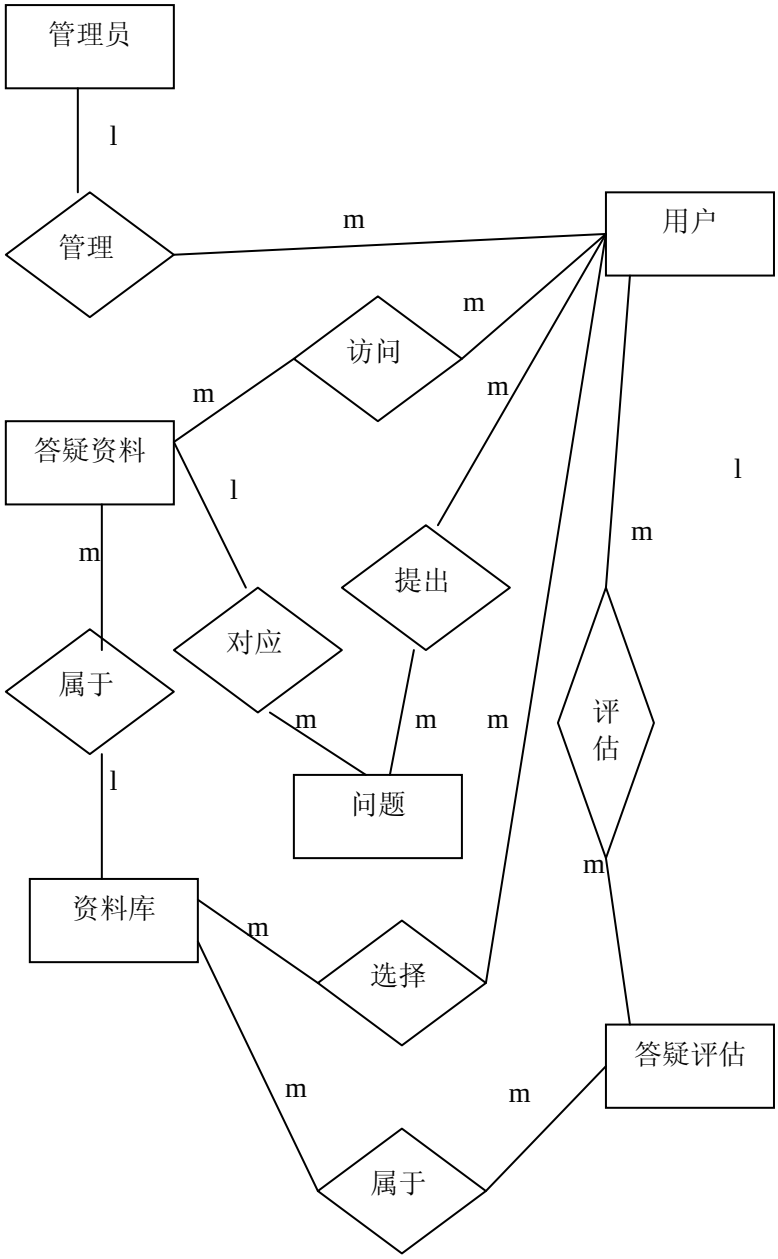


图 4-10 数据库实体联系图

4.5.2 创建数据库表

1.自动答疑信息表：包括问题标题、问题所属章节，专业关键字、一般关键字、解答内容、阅读次数、添加时间、提问人姓名等字段。如表 4-2 所示：

表 4-2 自动答疑信息表

字段	类型	整理	属性
id	Int(10)		自动编号

Question	Varchar(30)	utf8_general_ci	问题标题
Type	Int(10)	utf8_general_ci	问题所属的文章
Keywords	Varchar(30)	utf8_general_ci	专业关键词
Keyword	Varchar(30)	utf8_general_ci	一般关键词
Info	Text	utf8_general_ci	解答内容
Renum	Int(10)	utf8_general_ci	阅读次数
Date	Datatime	utf8_general_ci	添加时间
writer	Varcher(15)	utf8_general_ci	提问人的名字

2.异步答案信息表用于存放各个问题的基本信息，如问题的标题、作者、发表时间、以及回应数、点击数等。如表 4-3 所示：

表 4-3 异步答疑信息表

字段	类型	整理	属性
id	Int(10)		问题标识号
title	Varcher(30)	utf8_general_ci	问题的标题
writer	varchar	utf8_general_ci	提问作者
type	Int(10)	utf8_general_ci	所属章节的标识号
reply	Int(10)	utf8_general_ci	回答问题的标识号
info	text	utf8_general_ci	问题的内容
date	datetime	utf8_general_ci	问题发表的时间
ip	Varchar(15)	utf8_general_ci	提问者的 ip
replyx	Int(10)	utf8_general_ci	是否有答案
mark	Int(10)	utf8_general_ci	推荐文章
pic	longblob	utf8_general_ci	图片
newdate	datetime	utf8_general_ci	更新时间
renum	Int(10)	utf8_general_ci	点击数

3.同步答疑信息表主要包括有发送信息人、接受信息人、问题内容及提问时间。如表 4-4 所示：

表 4-4 同步答疑信息表

字段	类型	整理	属性
Child	Int(10)		自动编号

sender	varchar(30)	utf8_general_ci	发送信息人
receiver	varchar(30)	utf8_general_ci	接受信息人
content	text	utf8_general_ci	问题的内容
Time	datetime	utf8_general_ci	提问时间

4.关键词表主要存储关键字编号和内容。如表 4-5 所示：

表 4-5 问题答疑库关键词表

字段	类型	整理	属性
id	int		关键词编号
keyword	varchar	utf8_general_ci	关键词

4.5.3 数据库的存取访问设计

PHP 访问 MySQL 数据库的工作流程如图 4-11 所示：

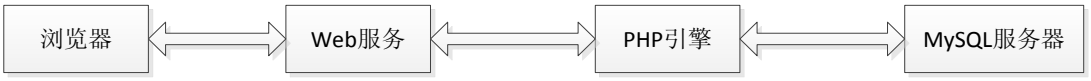


图 4-11 Web 数据库基本构架

1) 客户端通过 Web 浏览器对待定的 Web 页面发出 HTTP 请求。

2) Apache 服务器接收到用户的请求之后，获取到该文件，若为 PHP 文件，则把它传到 PHP 引擎执行处理。

3) PHP 引擎解析脚本，若文件中包含连接和查询数据库的命令，PHP 就建立通向 MySQL 数据库的连接，并发送适当的查询信息。

4) MySQL 服务器接收数据库的查询请求并处理，随后将结果返回给 PHP 引擎。

5) PHP 从数据库中获得数据后，依照程序的要求进行格式（HTML 格式）转换，然后再将具有 HTML 格式的文件交给 Apache 服务器。最后 Apache 服务器将其转送浏览器端以显示输出。

4.6 系统的实现

要达到的效果就是类似于站内搜索服务等，通过学生遇到的教学问题关键词匹配站内相关内容。以下是在 PHPMyFAQ 系统基础上优化后的 Test FAQ 教学辅助系统。

在此系统中可以通过关键词快速查找相关内容。

整体效果演示图如图 4-12：

图 4-12 效果演示图

普通用户不需要注册即可访问本答疑系统，可在系统内浏览所有静态页面。用户可以按章节查询：指用户访问系统后可按基本条件进行检索，可以分类查询：指用户进行更精确的按类别的检索。分类检索最多支持组配检索词的个数是三个，如：问题的学科领域、提问时间和所在章节等。教师则通过此界面来答疑：查看学生提出的问题，还能根据需要在三个不同类库的问题间进行转移。

提出问题效果图如图 4-13：

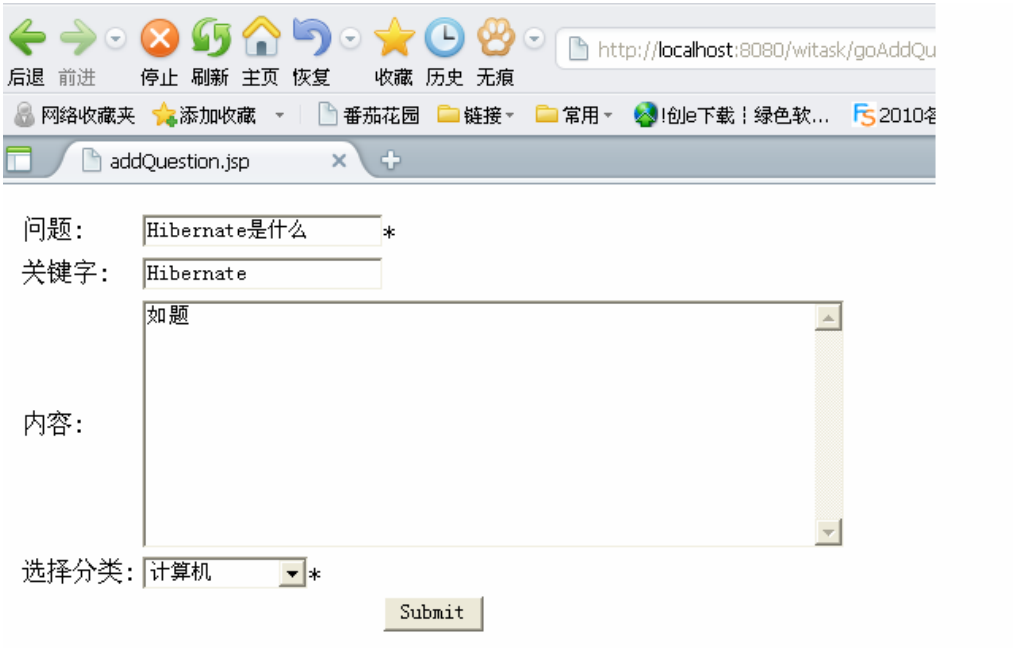


图 4-13 提问效果图

用户向系统输入拟提出的问题，同时结合具体规定来确定该问题的流程，问题被发往审核员。用户在新建问题时，首先通过自己的 ID 与密码进入系统，并选择提问管理模块。系统的“提问者”会自动添加用户的 ID，随后用户结合界面的

提示依次输入问题标题、问题主题，并对“问题类型”“秘密等级”“紧急程度”进行选择，然后添加问题内容，必要时上传问题附件，还可以对附件进行重新命名，并结合实际情况为问题添加备注。

用户点击“提交”按钮之后，系统会首先读取问题信息，并对问题进行必要的校验，如校验已获通过，则保存文件数据之后，询问用户是否现在开始假如提问正式流程，用户如果选择“是”，则依照系统预先定义的提问流程进行提问操作，此时的问题已经不可再更改。操作成功之后向用户返回提问结果，用户选择确认之后，提问步骤完成。核心代码为：

```
def_logic(self):
  askingquestions = 'askingquestions’in parameters and parameters['askingquestions']
or None
  askingtype : parameters['askingtype']
  askingpurpose = parameters['askingpurpose']
  askingstatus = parameters['askingstatus']
if askingname is None and askingtype ==“all" and askingpurpose ==“all" and
askingstatus==“all":
  assetlist = student.getassetlist（）
elseif askingquestions is not None:
if askingtype!=“all":
if askingpurpose!=“all":
if askingtype!=“all":
  assetlist=student.asking_factor_questionstypepurposestatus(askingquestions,
  askingtype,  askingpurpose,  askingstatus)
```

回答问题效果演示如图 4-14。

图 4-14 效果演示图

该界面罗列出新建问题的提问者、内容 、问题主题、问题类型等等属性，在回答员登陆系统之后，会在界面自动标记回答员的 ID，回答者可以选择指定具体的回答者，并在批示内容的文本框中添加答案。回答管理功能对问题的格式、内容等等方面进行回答。如果问题合格，则对其属性添加“合格”标识，如果问题不合格，则将其重新退回至提问者，问题需要重新修正之后，再进行回答。

如果回答通过，则用户需要指定问题的会打人，系统对问题进行自动校验，确认无误则对问题及其答案进行保存。接着读取预先设定的问题回答流程，并向用户显示回答结果。回答通过之后的问题，其属性带有“已回答”的标识。

问题搜索效果如图 4-15。

图 4-15 效果演示图

用户以各种方式实现查询，包括输入单个条件的直接查询，输入多个条件的组合查询以及输入条件的模糊查询等，用户查询到问题之后，能够结合自己的权限对该问题做出相应的处理。

用户在查询系统相关内容时，通过输入查询条件，或者不同类型查询条件的组合，然后选择查询结果的排序方式，并进行 SQL 语句语法校验，最后系统返回与条件相匹配的查询结果。核心代码为：

```
for (int i = 0; i < ids.length; i++) {
    TplQuestionsAskingItem tplQuestionsAskingItem = (TplQuestionsAskingItem)
    questionsaskingitemservice.queryTplQuestionsAskingItemById(Long.valueOf(ids[i
    ]));
    Long putaskingId = tplQuestionsAskingItem.getPutaskingId();
    String packId = tplQuestionsAskingItem.getPackId();
    if(packId.startsWith("WW")){
        throw new Exception("msg:答案不存在，请核实!");
    }
    if (null != putaskingId) {
        context.getRequest().setAttribute("msg","已回答，请核实");
        return this.queryTplQuestionsAskingMgrListByContext(context);
    }
    if(procFlagService.isTubeProcFlag(new Long(ids[i]),
    "packAsking")){
        throw new Exception("msg:请重新选择!");
    }
```

}

4.7 本章小结

本章主要阐述了PHPMyFAQ系统的模块分析和实现的过程。首先在Test FAQ模块中采取自然语言的提问方式，然后将提取的关键字与问题答疑库做匹配，得到用户所提问题的答案，如果用户对该答案不满意，可以进一步将问题提交到异步答疑中。其中的关键技术是正向最大匹配的中文分词方法，用于处理提交的文本。由于Test FAQ模块只是网上答疑系统的一部分，就没有在此介绍同步和异步答疑系统。

第五章 系统测试

一个成功的项目不仅仅是完成一个步骤而已，还需要测试维护等工作，在完成PHPMyFAQ系统分析工作以后还对系统进行了优化和测试等工作，从而让系统更加完善。

5.1 Apache 等服务的监测

PHPMyFAQ系统是Web应用，所以需要保证应用一直处于开放状态，我们需要确保我们系统的 Apache、MySQL 等服务的正常运行，需要对 Apache 等的工作状态经行维护以保证系统的完整性。应用架构与Linux之上，我们采用Linux下的系统调度进程Cron去维护apache的运行状态。Cron是一个Linux下的定时执行工具，可以在无需人工干预的情况下运行作业。首先通过 ps －ef 列出当前所有进程 然后在列表中匹配apache 进一步匹配httpd的服务项 最后通过wc -l列出有符合规则的数目并赋值给i。如果大于等于1则确认apache在运行，如果小于则启动apache。

代码如下：

```
#/bin/sh
i=`ps –ef |grep '/apache/'|grep httpd|wc -l`
if [ "$i" -ge 1 ]; then
echo "The process already run\n";
else
echo "Start the process now:\n"
/data1/apache/bin/apachectl start
fi
```

然后通过 cron 去定时执行而维护 apache 的状态。同理监控 Memcached 也是用类似方法，一下如监控 Memcached 的 shell 脚本。

```
#!/bin/sh
#check memcache process and restart if down
mm_bin="/usr/local/bin/memcached"
mm_log="/home/xxx/memcached_check.log"
mm_ports=("11211" "11212")
```

```
mm_param=("-d -m 20480 -p 11211 -u www" "-d -m 256 -p 11212 -u www")
mm_count=${#mm_ports[@]}
t=$(date -d "today" +"%Y-%m-%d %H:%M:%S")
i=0
while [ $i -lt $mm_count ]
do
mm_exists=`ps -ef|grep "memcached"|grep "${mm_ports[$i]}"|grep -v grep|wc -l`
if [ "$mm_exists" == "0" ]; then
${mm_bin} ${mm_param[$i]} 2>&1 > /dev/null &
echo "${t} : ${mm_bin} ${mm_param[$i]}" >> ${mm_log}
fi
let i++
done
```

5.2 压力测试

操作系统知识库系统不仅仅需要满足实时性，还需要保证在用户过多的时候也能够保证良好的运行，我们采用 Apache 中的测试工作 AB 测试工具，其设计意图是描绘当前所安装的 Apache 的执行性能，主要是显示你安装的 Apache 每秒可以处理多少个请求。AB 的使用有很多方式我们可以做一个最简单的测试在命令行输入。

```
./ab -n number_of_total_requests \
-c number_of_simultaneous_requests \
http://localhost/test.php
```

例如：

```
./ab -n 1000 -c 50 http://www.localhost.com/test.php
```

AB 将同时向 http://www. localhost.com/test.php 发出 50 个并发请求，共发出 1000 次。

测试结果是这样的：

```
Server Software: Apache/2.0.16
Server Hostname: localhost
Server Port: 80
Document Path: /test.php
Document Length: 1311 bytes
Concurrency Level: 50
```

```
Time taken for tests: 8.794 seconds
Complete requests: 1000
Failed requests: 0
Total transferred: 1754000 bytes
HTML transferred: 1311000 bytes
Requests per second: 113.71
Transfer rate: 199.45 kb/s received
Connection Times (ms)
min avg max
Connect: 0 0 5
Processing: 111 427 550
Total: 111 427 555
```

test.php 每秒钟可以处理的请求数为 113.71 个。将请求数增加，看看服务器能否处理更大的压力。你也需要调节 Apache 的 MaxClients， MaxThreadsPerChild 等参数，基于你的 httpd.conf 中的 MPM 模块选择。

压力测试是为了更好的得知系统的薄弱环节在得知后并做进一步的优化提供系统稳定性，当然处理 AB 测试工具之外还有很多其他的测试方法和工作。

5.3 安全测试

安全测试是最重要的测试，如果系统没有安全性，也就没有了他存在的价值。安全测试最要根据当前流行的攻击手段去模拟攻击进行测试。比如 SQL 注入，XSS 跨站脚本攻击，CSRF 跨站请求伪造等攻击手段。不仅仅测试需要注意安全，在开发中更需要注意安全，因为大部分漏洞都是开发中考虑不周造成的，养成一个良好的编程习惯。首先需要认识到一下几点。

- 任何用户的输入都是不安全的
- 过滤数据远远没有限制输入安全
- 最少的服务+最小的权限=最大的安全
- 头脑清晰，避免逻辑漏洞

安全测试主要完成 MySQL 注入测试，SQL 注入是参数传递过滤不严时用参数组合成正确的 SQL 语句来执行特定操作的技术。简单的理解为：涉及进入 SQL 语句操作的变量没有过滤输入与转义输出。

注入示例：or 语句+注释：

```
Select * from user where uname = ‘$a’ and password = ‘$b’
    $a = 1 ' or 1=1 --
```

```
$a = 1 ' or 1=1 /*
Select * from user where uname = '1 ' or 1=1
union 查询
select A，B，C from user where type = $id
$id = A union select D，E，F from admin_user where type = B
select A，B，C from user where type = A union select D，E，F from admin_user where type = B
```

当测试去 MySQL 语句存在漏洞时需要完善，主要有以下方案应对 SQL 注入攻击：

- 来自用户的数据，首先要进行过滤处理
- 白名单，黑名单，数据类型，数据格式…

OR， SELECT， UPDATE， UNION…

- 来自用户的数据，在进入数据库前作转义处理

Select * from user where uname = ‘1\ ' or 1=1

- 使用单引号限制或整数处理

SELECT * FROM TABLE WHERE cid = ‘123 and ord(mid(version()，1，1))>51’

SELECT * FROM TABLE WHERE cid = 123

除了 MySQL 的一些问题之外还有很多安全问题，设计到 PHP 的配置例如需要设计 PHP 配置文件中的 display_error 为 off，这样目的是防止错误信息泄漏。我们一下介绍一下在安全测试需要用的到工具以及技巧。

首先我们可以通过文件内容分析的工具—grep 进行对代码的分析。当然还用到了 PHP Xcode，WebInspect ，Acunetix Web Vulnerability Scanner。

5.4 浏览器兼容性测试

兼容性测试是将验证软件与其所依赖的环境的依赖程度，包括对硬件的依赖程度，对平台软件，其它软件的依赖程度，来检查程序能正常的运行的测试。现在浏览器越来越多，从 IE 浏览器以前完全的统治着人们使用浏览器上网的行为习惯，到目前随着 Google，Firefox 等其他浏览器的迅速窜起，虽然给一些厌倦了 IE 浏览器的人们提供了更多不同的选择，同时也给用户越来越多的 DIY 使用体验，但是却给页面重构工作者的工作量越来越繁重。由于浏览器大战等历史原因，Web 前端存在着很多的兼容性问题。浏览器兼容性测试主要是查看在不同的浏览

器下页面出现的不同效果，目前还没有太好的解决方案，但可以解决工具更加方便的经行测试。比如 fiexbug、web develop 等工具进行测试与分析。当然测试工作还有很多，目前简单介绍了在 web 应用中比较重要的测试。

5.5 本章小结

本章主要阐述了对 PHPMyFAQ 系统的测试工作，包括对 Apache 等服务的检测、压力测试 、安全测试和浏览器兼容性测试等内容，对本文工作的有效性进行了验证。

第六章 总结和展望

6.1 总结

随着网络不断拓展，网上教学也正逐步推广开来，因此网上答疑显得必要而迫切。网上答疑系统设计是网络教学平台设计中的主要部分，是实施网络教学和学生网上自主学习的主要组成部分。网络教学是目前高校中对传统教育很好的辅助，不仅为学生自主学习提供了很好的环境，也大大提高了教学质量，促进师生之间的沟通。因此，对于它的设计要务必考虑方案精良、功能全面、性能高效，力图使之发挥更大作用。现将本文的研究工作总结如下：

1．本系统的特点是采用关键词搜索的方式充分发挥网络优势，为教学辅导在时间和空间上提供更大的灵活性，学生可根据自己的安排机动掌握参与辅导的方式，教师也可获得更多关于学生和课程的反馈信息，为教学提供有益参考。

2．自动答疑系统支持用户以自然语言和输入关键字两种方式提问，其中自然语言方式支持中英文混合输入，采用正向最大匹配法对问题进行切分，提取关键词，最终将系统中存在的问题答案提供给用户。如有问题在自动答疑系统中未能解答，系统会提示将该问题转入异步答疑系统。

3．异步答疑模块提供了浏览、搜索、提问、解答、管理等功能，学生除了可以在自动答疑模块中提问，还可以在此模块中提问，教师主要在这个模块中解答来自自动答疑模块和异步答疑模块的问题，并且为教师提供了管理的功能。

4．本系统以课程为单位和基础，较好的实现了通用性设计，并且具有较好的可扩展性，使得本网络课程网站可以推广到其它的课程。同时，课程中采取了按章节归类。

5．同步答疑模块被目前很多答疑系统所忽视。即使答疑系统具有很高的智能性，也不可能达到人的智能性，只有教师才能对学生的提问理解准确，只有教师才能深刻把握存在于学生问题背后的逻辑错误或创造火花，只有教师才能有针对性的回答学生的提问。

6.2 展望

传统的教学手段必须引入新的模式，才能够满足当前网络技术迅猛发展带来的海量内容信息的教学需求，为适应当今社会发展对高素质创新型人才的需要，创造一个在教师指导下的学生自主式学习环境，是当前的研究热点之一。网络教

育形式融合了信息技术与教育，使得学习机会更便捷，学习环境更多样，学习资源更丰富，学习活动更自主、个性化，增强了教育的适应性，为教育注入了新的活力，因此近年来备受关注。

致谢

光阴荏苒，硕士研究生的学习即将结束，三年的学习生活使我受益匪浅。经历大半年时间的磨砺，硕士毕业论文终于完稿，回首大半年来收集、整理、思索、停滞、修改直至最终完成的过程，我得到了许多的关怀和帮助，现在要向他们表达我最诚挚的谢意。

首先，我要深深感谢我的导师吴跃老师。吴老师为人谦和，平易近人。在论文的选题、搜集资料和写作阶段，吴老师都倾注了极大的关怀和鼓励。在论文的写作过程中，每当我有所疑问，吴老师总会放下繁忙的工作，不厌其烦地指点我；在我初稿完成之后，吴老师又在百忙之中抽出空来对我的论文认真的批改，字字句句把关，提出许多中肯的指导意见，使我在研究和写作过程中不致迷失方向。他严谨的治学之风和对事业的不懈追求将激励和影响我的一生，我更将永远铭记他对我的教诲和关心。在此，我谨向吴老师致以我最衷心的感谢。

同时也感谢各位同事，他们非常热情地帮助我收集第一手语料，感谢他们在本文调研工作期间对我所提供的大力支持与帮助。我对我的家人说声谢谢，谢谢我的爱人和我的父母，他们给了我最大的鼓励与最朴实的帮助。

最后，我要感谢各位论文评审专家和答辩专家，他们给了我一个审视几年来学习成果的机会，让我能够明确今后的发展方向，他们对我的帮助是一笔无价的财富。我将在今后的工作、学习中加倍努力，以期能够取得更多成果回报他们、回报社会。再次感谢他们，祝他们一生幸福、安康!

参考文献

[1]蔡冠群，张业睿，袁晓斌等.构筑基于 Web 的远程答疑系统[J].信息技术教育，2006 (03)：75-76.

[2] J.Janes，D.Carter，P.Memmott.Digital Reference Service in Academic Libraries. Reference and User Services Quarterly[J]，2011，39(2):145-150..

[3]张小艳.中文自动的研究与实现[J].微计算机信息，2007 (36)：208-210.

[4]孙宏才.网络层次分析法与决策科学[M].北京:国防工业出版社，2011，1.

[5]周江卫.数字学位论文系统的研究与开发[D].西安电子科技大学，2003.

[6](美)勒道夫著，陈浩，胡丹，徐景译.PHP 程序设计(第 2 版)[M].电子工业出版社，2007.

[7](美)Elliott White III，(美)Jonathan D.Eisenhamer，王军译.PHP 5 in Practice[M]。中文版人民邮电出版社，2007.

[8]詹海菊.基于 PHP 技术的网站设计[J].科技信息(学术研究)，2008 (22)：174-176.

[9]Luke Welling，Laura Thomson 著.PHP and MySQL Web Development third Edition[M]机械工业出版社，2005.

[10]袁萌.MySQL 让自由力量放光彩[J].信息系统工程，2007 (5) ： 82-85.

[11]Mandar Chitnis，Pravin Tiwari， Lakshmi Ananthamurthy，林艳芹.UML 工具简介[J].软件世界，2007 (22) ： 44-45

[12]卢玉婕.Web 数据库技术研究与应用[D].北京交通大学，2004.

[13]刘件，魏程.中文分词算法研究[J].微计算机应用，2008 (08) ：11-16.

[14]徐飞，孙劲光.中文分词切分技术研究[J].计算机工程与科学 2008 (05)：126-128.

[15]郭晓燕，张博锋等.自动答疑中问句相关度算法研究及系统实现[J].计算机应用，2005 (2).

[16]何苹，王碗芜.自然语言检索中的中文分词技术研究进展及应用[J].情报科学，2008 (05)：787-791.

[17]胡锡衡.正向最大匹配法在中文分词技术中的应用[J].鞍山师范学院学报，2008 (02)：42-45.

[18]贺桂英，李拥军.远程教学智能问题流模型的研究和实现.电化教育研究，2002(08)

[19]吉逸，虞维平.远程教育教学支撑平台的研究与实现.南京师大学，2002

[20]杨兴凯，秦玉平.用 Servlet 实现自动. 锦州师范学院学报(自然科学版)2002(01)

[21]贺桂英，李拥军.远程教学智能问题流模型的研究和实现.电化教育研究 2002(08)

[22]李康.远程教育理论中几个基本概念的探讨.中国远程教育，2003(05)

[23]柳泉波. 智能的设计与实现. 中国远程教育，2008(08)

[24]李伟华. 网上教学系统中知能答疑专家系统的设计，2004

[25]张伟元. 网上学习环境评价模型、指标体系、及测评量表的设计与开发[J]. 中国电化教育，

2004(07):29-33

[26]何克抗. 计算机辅助教育[M]. 北京：教育出版社，1997

[27]刘笑冬.中国高校图书馆网上参考咨询服务研究.湖北函授大学学报[J]，2012，23(5):34-35.

[28]郑耿忠.自动分词算法在智能中的应用研究. 计算机工程与设计，2007(09)

[29]陈展荣，曾毅平. Web 汉语料的智能抽取与词汇切分[J]. 计算机工程与设计，2005(06)

[30]赵成龙，薛欣. 基于 WEB 的智能的设计与实现[J]. 安阳师范学院学报，2004(02) 69

[31] 刘笑冬.中国高校图书馆网上参考咨询服务研究.湖北函授大学学报[J]，2010，23(5):34-35.

[32] 狄德罗.大英百科全书[M].美国:不列颠百科全书公司出版，2002，961.

[33] 中国大百科全书总编辑委员会.中国大百科全书[M].北京:中国大百科出版社，1993.

[34] Jo Bell Whitlatch. Evaluating Reference Services:A Practical Guide[M].Chicago:American Library Association，January 2010.

[35] William A.Katz. Introduction to reference work，2V Reverence Services and Reference Processes[M].New York:McGraw Hill，2010.

[36] 初景利.图书馆数字参考咨询的理论与实践研究[D].北京:中国科学院研究生院（文献情报中心），2012 年.

[37] 向上.图书馆数字参考咨询系统研究与技术实现[D].成都:四川大学，2012.

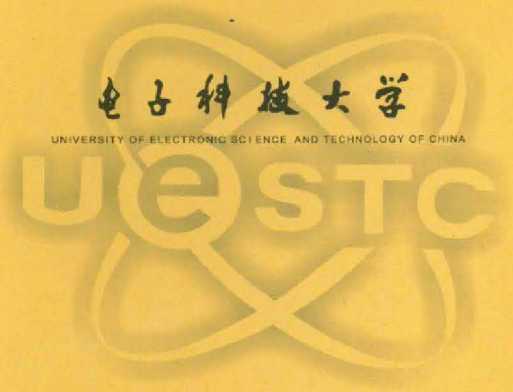
[38] 孔慧.参考馆员与学科馆员.医学信息学杂志[J]，2011(1):46-47.

[39] 詹德优.信息咨询理论与方法[M].武汉:武汉大学出版社，2012

[40] 詹德优.关于新中国参考咨询研究的统计与分析.图书与情报[J]，2010(1):20-25

[41] 詹德优.20 世纪中国参考咨询服务:发展历程，成就与局限.高校图书馆工作[J]，2010，20(77):1-7，28

[42] 胡典慧.基于 PHP 的网上辅助教学系统设计与实现[D].东北大学, 2010



专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE