

专业设计评分表

学生姓名: 王帆 学号: 20152180 日期: _____

专业: 计算机科学与技术 指导教师: 李玉坤

考核项目	考核点	占 分 比例	参考评价标准	得分
平时成绩	日常表现	20	积极认真完成阶段性任务 18-20 分 较为积极认真完成各阶段任务 12-16 分 不能认真完成阶段项任务 0-10 分	
	学习能力	10	阅读文献及独立解决问题能力强 8-10 分 阅读文献及独立解决问题能力一般 6-7 分 阅读文献及独立解决问题能力弱 1-6 分	
	协作与沟通能力	10	设计中表现出较强的协作与沟通能力 8-10 分 协作与沟通能力一般 6-7 分 写作与沟通能力弱 <6 分	
设计方案	创新性	5	问题新颖或解决方法具创新性 4-5 分 创新性不显著 <4 分	
	设计方案复杂性	10	数据结构及设计算法具有一定复杂性 8-10 分 数据结构或设计算法具有一定复杂性 6-8 分 数据结构和设计算法均过于简单 <6 分	
	设计方案合理性	10	设计方案合理、可行 6-10 分 设计方案可行性方面有问题 <6 分	
	设计方案综合性 (包括对社会、环境、文化、安全、法律等的影响)	10	考虑了设计方案对社会、法律、安全、环境、文化等的影响 8-10 分 部分考虑了对上述因素的影响 1-7 分 没有考虑对上述因素的影响 0 分	
	设计方案的实现	10	完整进行了程序设计与实现 8-10 分 实现了部分功能 1-7 分 没有进行系统实现 0 分	
	设计报告质量(包括文档规范性、语言逻辑性等)	15	格式规范、语言通顺 12-15 分 格式基本规范、语言基本通顺 9-11 分 格式、语言问题较多 1-8 分	
考核最终成绩				

指导教师签字: _____

专业设计报告

学号：20152180

姓名：王帆

专业：计算机科学与技术

指导教师：李玉坤

设计题目：高职院校志愿信息系统

天津理工大学计算机学院

1. 问题描述

1.1 课题背景及介绍

高考是人生较为重要的考试之一。然而由于高校信息较为繁杂，数据不易查询，传统的报考指南过于冗繁，且不利于检索。传统模式下的院校专业信息仅能通过纸质资料或各院校及考试院发布的零散数据中获取，费事费力。在信息化的今天，如何提高信息检索的效率是一个非常具有研究价值的课题。当前市面上已经存在不少关于本科院校招生信息系统产品，而高职高专院校较少，仍处于空白。同时，在用户体验至上的移动互联网时代，如何提升用户的使用体验也是一个十分重要的课题。这种情况下，面向高职高专考生与家长的高职院校志愿信息系统应运而生。

1.2 课题研究目的及应用

高职院校志愿信息系统，可以为考生和家长提供较为友好的使用体验，以及较为便捷的信息检索方式。而通过高职院校志愿信息系统的设计，可以实现院校数据的聚合与专业信息的整理，帮助考生做出决策，考上理想的大学。同时，针对分数较低但有意愿参与专升本考生的考生，系统提供了专业对接查询功能，方便考生选择专科专业的同时，了解该专业未来发展方向，及专升本可报考的专业信息等，为考生与家长查询报考志愿提供一定的参考与帮助。

高职院校志愿信息系统，是考生与家长志愿报考与院校选择的工具与助手。结合用户需求和当今发展趋势，开发具有移动适配性与便捷性的高考信息检索系统是十分必要的。

1.3 设计开发及运行的软硬件环境

1.3.1 开发环境

本系统采用的开发平台为 Windows 10，使用 VBScript 语言并基于 ASP 开发，采用 IIS 10.0 作为 Web 服务器。数据库方面采用 MariaDB 10.3，并使用 Navicat 作为数据库管理系统。前端部分使用 CSS / HTML 构建，具有较好的扩展性与普适性。

MariaDB 是一个采用 Maria 存储引擎的 MySQL 分支版本，相对于 MySQL 5.6 来说，在性能、功能、管理、NoSQL 扩展方面，MariaDB 包含了更丰富的特性。比如微秒的支持、线程池、子查询优化、组提交、进度报告等。在兼容性方面，MariaDB 能够提供较好的向下兼容 MySQL 方法。

1.3.2 运行环境

本系统可在 Windows 10 等主流操作系统，基于 Chrome/Firefox 等主流浏览器，在 1920×1080 分辨率下运行。

2. 文献阅读情况

随着互联网的迅速普及，人类迅速进入了信息时代，在这个信息社会里，谁掌握了信息，谁就在竞争中处于有利地位。在这种形式下，如何能在大量纷繁复杂的信息中找到需要的有用信息，就成为了人们研究的热点。^[1]

2005 年 8 月 IBM 推出了开源的 UIMA，这是一项能够为非结构化的文本、图像、音频和视频等信息提供分析功能的技术规范，可以实现对非结构化信息的内容分析和数据挖掘，这一规范一经推出就得到 Attensity、CleaForest、SAS、Spss 等 16 家全球知名软件公司的支持，此后经过 3 年多的宣传和推广，2009 年 3 月 19 日国际开放标准联盟批准 UIMA 1.0 版本为 OASIS 标准^[2]，这表明 UIMA 将对全球范围内的内容分析和数据挖掘起着规范作用，并为非结构化信息的分析、抽取、整合和利用提供一个通用的平台。企业利用 UIMA 可以把非结构化的文本、图像、音频和视频等信息转换成结构化的信息，在此基础上进行深层次的开发和利用。^[3]

由于研究中文信息抽取起步相对较晚，加之中文所特有的特殊性，构建基于中文的信息抽取系统要比构建西文（例如：英文）信息抽取系统要困难很多。目前，我国国内的一些搞笑和研究机构在进行相关研究，主要有北京大学、上海交通大学、哈尔滨工业大学和中科院等，研究的重点主要是简单抽取任务和识别中文命名实体，目前还在探索中的是设计能够对中文信息进行完整抽取的系统。^[4]

此外，对于相关信息的相似度判定也是一个重要的课题。目前国内外对文本相似度的研究方法主要分为基于统计的方法和基于语义分析的方法。^[5]基于统计的文本相似度量方法是将文本表示为特征词集合，将特征词作为文本的基本元素，建立特征词向量空间，通过计算特征词向量空间之间的相似度来表征文本之间的相似度。这种方法需要大规模的语料库，忽略了文本中的语法和组织结构等信息，也忽略了文本中的词语语义信息，且文本表示模型高维稀疏，计算效率低下。基于语义分析的文本相似度量方法则是利用特定领域的知识库来构建文本特征词之间的语义关系。这种方法不需要语料库，但建立完备的知识库则比较费时费力。因此将二者相互结合，既能有效降低文本表示模型的维度，又能结合词项语义信息进行相似度计算，可以高效快速地计算出两个文本在语义层次上的相似度，并且能够在较为广泛的应用领域内使用。^[6]

3. 设计约束及解决策略

3.1 技术可行性

技术可行性是指在特定条件下，分析技术资源的可用性和这些技术资源用于解决信息系统问题的可能性和现实性。

本系统采用 ASP 实现 B/S 架构的信息系统，VBScript 学习成本低，实现简

单。同时，对于数据处理方面，我们使用 Java 语言作为数据处理语言，实现数据规范化、数据入库、数据维护等操作。

爬虫框架方面，我们采用基于 Java 的 Jsoup HTML 解析器。Jsoup 解析 HTML 过程与 jquery 选择器的语法非常相似，并且非常灵活容易使用以获得所需的结果。使用 Jsoup，我们可以从 URL，文件或字符串中刮取并解析 HTML，查找和提取数据，使用 DOM 遍历或 CSS 选择器，操纵 HTML 元素，属性和文本，并输出整洁的 HTML。

前端展示层方面，我们采用当前流行的 Bootstrap 前端开发框架。Bootstrap 是美国 Twitter 公司的设计师 Mark Otto 和 Jacob Thornton 合作基于 HTML、CSS、JavaScript 开发的简洁、直观、强悍的前端开发框架，使得 Web 开发更加快捷。Bootstrap 提供了优雅的 HTML 和 CSS 规范，能够实现较为美观且实用的前端页面。

3.2 经济可行性

经济可行性是分析信息系统项目所需要的花费和项目开发成功之后所能带来的经济效益。分析信息系统的经济可行性，就是分析该系统是否值得开发。一个信息系统的总成本包括开发总费用和运行管理总费用，信息系统总效益包括直接经济效益和间接社会效益。

3.3 可能导致的社会与法律问题

本课题研究的是信息系统集成的设计，因此需要从许多数据源获取相关数据信息。这些数据源，有些是来自于国家部委下发文件，有些是部分现有信息系统已有数据。作为研究使用的相关数据，我们将予以标识，说明版权来源，以规避侵权问题。

4. 设计方案

4.1 概要设计方案

4.1.1 E-R 图

在本系统中涉及四个实体：院校、高职专业、高职专业大类、本科专业。另外包括院校与高职专业的对应关系。其中院校分为高职高专院校，以及部分开设高职高专课程的普通本科院校。对院校-专业对应表进行内部编码处理，一方面提供了未来扩展的接口，便于分数信息的管理实现，另一方面也满足了通过索引对数据库进行快速查询的需求。高职专业大类、高职专业和本科专业之间是一对多关系，其中，每个专业大类至少包括一个专业，每个高职专业可对应多个本科专业。

根据用户需求和高考信息检索系统的设计规范，我们这样设计院校、高职专业、高职专业大类和本科专业等实体的属性：“院校”实体具有院校代码、院校名称、院校排名、省内排名、评分等级、学科类别、院校类型、主管部门、所在省区、院校简介、收费细则、官方网站、曾用名称这十三项属性，其中院校代码为主属性；“高职专业”实体具有专业代码、专业大类代码、专业名称、专业类型、培养目标、专业课程、专业技能、工作方向等属性，其中专业代码为主属性。专业大类代码和专业大类名称是“专业大类”实体的两个属性，其中专业大类代码为主属性。“高职专业对应本科专业”实体具有内部编码、专业大类代码、专业代码、专业方向、对应本科专业名称等属性，其中内部编码为主属性。实体关系的详细模式如图 4.1：

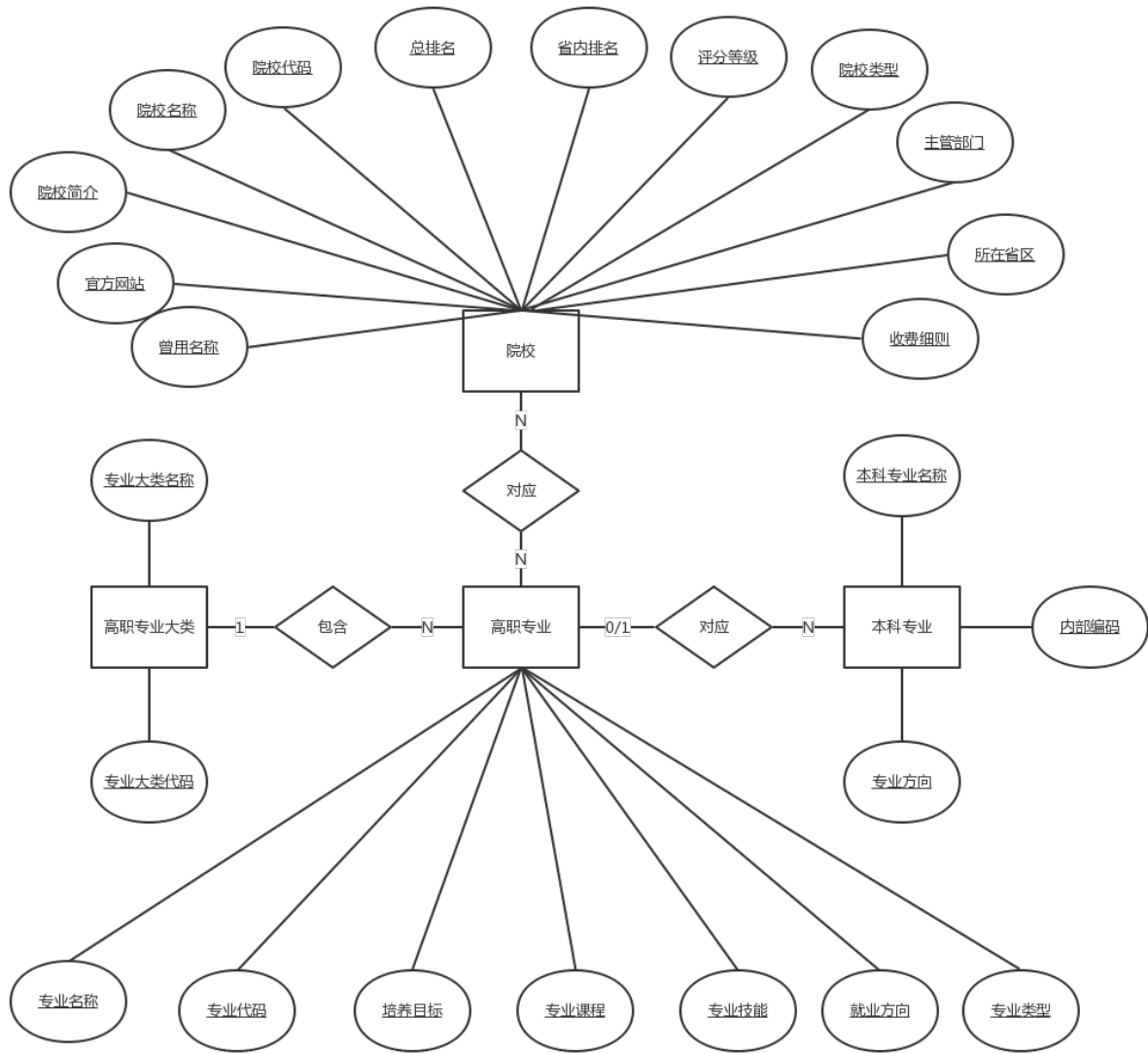


图4.1 实体关系图

本系统共使用了 5 张数据表，分别控制院校、专业、专业大类、院校排名、

高职专业与本科专业对应关系，其逻辑结构模型如图 4.2 所示：

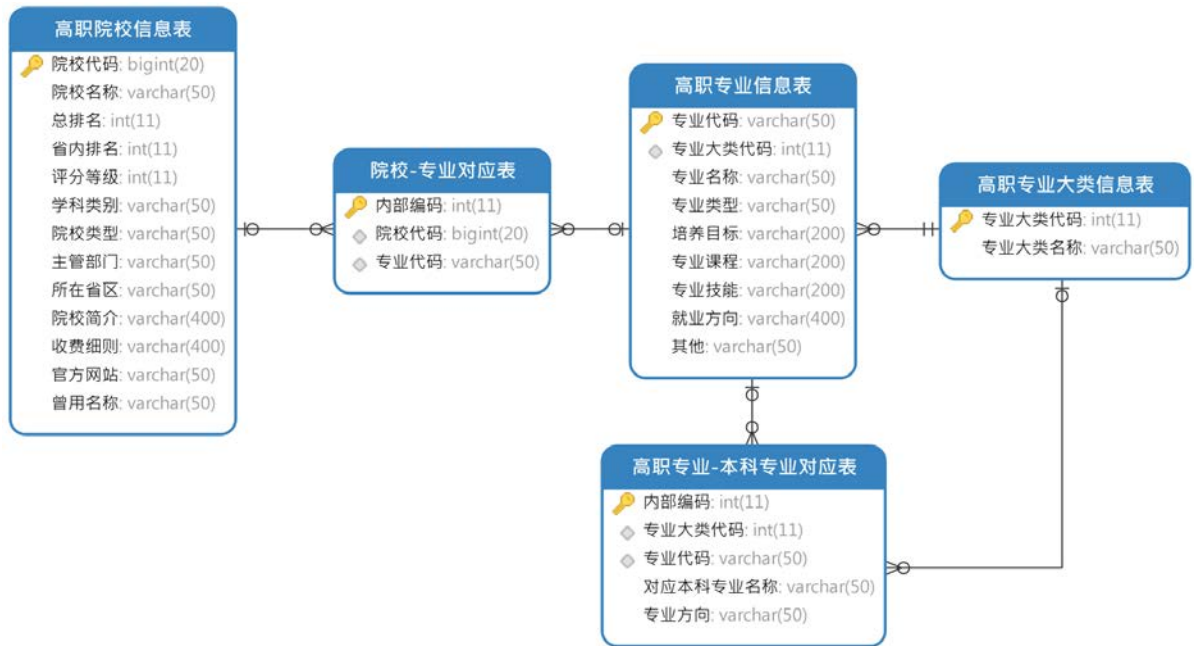


图4.2 数据库逻辑结构模型

关于以上实体和关系的属性以及对属性的必要说明和约束关系，如表 4.1 所示：

表4.1 各实体属性说明

实体	属性	说明	类型	约束
院校	school_code	院校代码	big integer	主键
	school_name	院校名称	varchar(50)	
	school_rank	院校排名	integer	
	province_rank	省内排名	integer	
	school_grade	评分等级	integer	
	subject_type	学科类别	varchar(50)	
	school_type	院校类型	varchar(50)	
	head_name	主管部门	varchar(50)	
	province	所在省区	varchar(50)	
	info	院校简介	varchar(400)	
	fee	收费细则	varchar(400)	
	website	官方网站	varchar(50)	
	formal_name	曾用名称	varchar(50)	
专业	major_code	专业代码	varchar(50)	主键
	major_class_code	专业大类代码	integer	外键
	major_name	专业名称	varchar(50)	
	major_type	专业类型	varchar(50)	
	major_goal	培养目标	varchar(200)	
	major_course	专业课程	varchar(200)	

	major_skill	专业技能	varchar(200)	
	work_direction	工作方向	varchar(200)	
	others	其他	varchar(200)	
专业大类	major_class_code	专业大类代码	integer	主键
	major_class_name	专业大类 名称	varchar(50)	
高职对应 本科专业	id	内部编码	integer	主键
	major_class_code	专业大类代码	integer	外键
	major_code	专业代码	varchar(50)	外键
	major_direction	专业方向	varchar(50)	
	to_university_major	对应本科专业名称	varchar(50)	

4.1.2 功能模型

基于对用户需求的分析，结合设计规范，用户与系统之间的数据交换只有信息查询（数据流入）和结果显示（数据流出）两部分。根据以上分析，我们设计出数据流图，如图 4.3 所示。

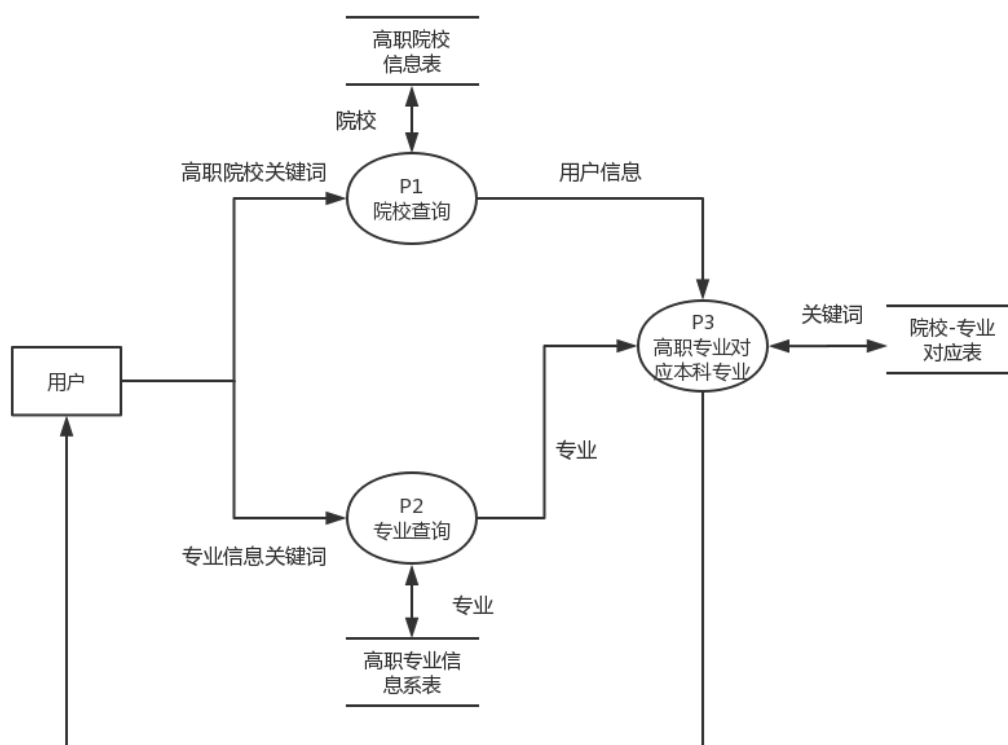


图4.3 顶层数据流图

实体-关系图当中的专业实体，通过对用户输入关键信息与类型不同可以分别实现对不同属性的针对性查询，包括大类关键字、专业关键字、代码关键字等。结合以上内容，下一步我们对上图做进一步的补充，设计出功能级数据流图，如图 2.4 所示。

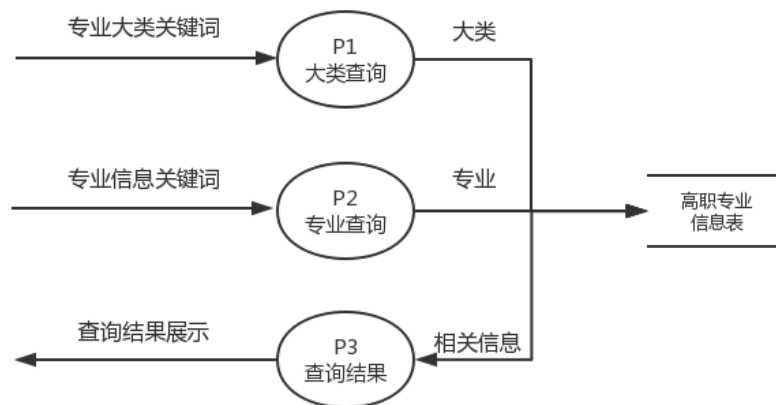


图4.4 功能级数据流图

4.2 详细设计方案

4.2.1 功能分析

前述的功能需求共有四个部分，分别是院校信息查询，专业信息查询，院校排名查询，高职专业与本科专业匹配筛选。结合系统模型对以上功能进行分析，我们可以有如下发现：

一、院校信息查询和专业信息查询是系统的两大最核心功能。

二、高职专业与本科专业匹配筛选是专业信息查询功能的一个部分，也可以看作专业信息查询功能的衍生功能，是对高职专业匹配本科专业后进行输出的功能。

三、院校信息查询是与专业信息查询相并列的一个功能，主要是对高校基础信息的检索。

四、院校排名查询是院校信息查询的一个子功能。通过对数据进行编码排序，可以导出其对应的院校排名信息，并链接至对应院校详情页面。

根据以上分析，我们就可以知道前述四大功能之间的区别和联系。下一步，我们就需要以功能及其相互关系为基础，进行模块分解和结构设计。

4.2.2 模块分解和结构设计

我们将整个系统的功能划分为两个主要部分：院校信息与专业信息。

院校信息部分，包含院校详细信息和院校排名信息两个模块，实现的是对院校数据库针对性检索功能，并通过字段“排名”进行联系。

专业信息部分，包含专业大类信息，专业详细信息和高职专业对应本科专业信息三个模块。高职专业对应本科专业信息是针对给定高职专业进行本科专业筛选的功能，未来增加普通本科院校信息后，可同本科信息进行链接显示。

结合上述模块分解，我们设计出系统结构，如图 4.5 所示。

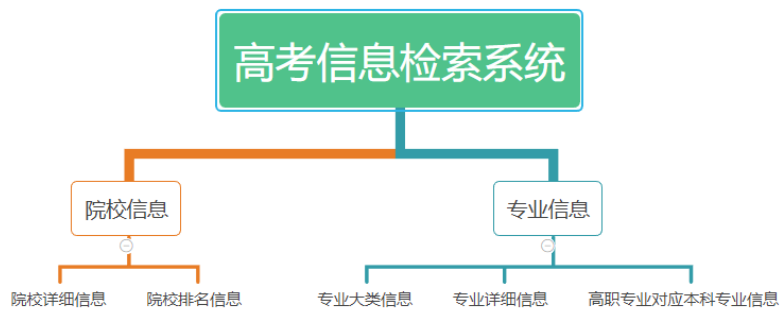


图4.5 系统总体结构图

4.2.3 核心算法设计

核心算法方面，主要是关于数据抽取的算法设计与实现，包括结构化数据抽取与非结构化数据抽取两部分。

结构化数据抽取方面，采用使用 jsoup 对 HTML DOM 进行解析与遍历，使用 `getElementById()` 方法遍历一个 Document 对象，并使用 `Element.select` 抽取特定标签/类/id 的数据，最后在遍历过程中将数据进行组织，从而实现结构化数据的抽取。算法流程图如图 4.6 所示。

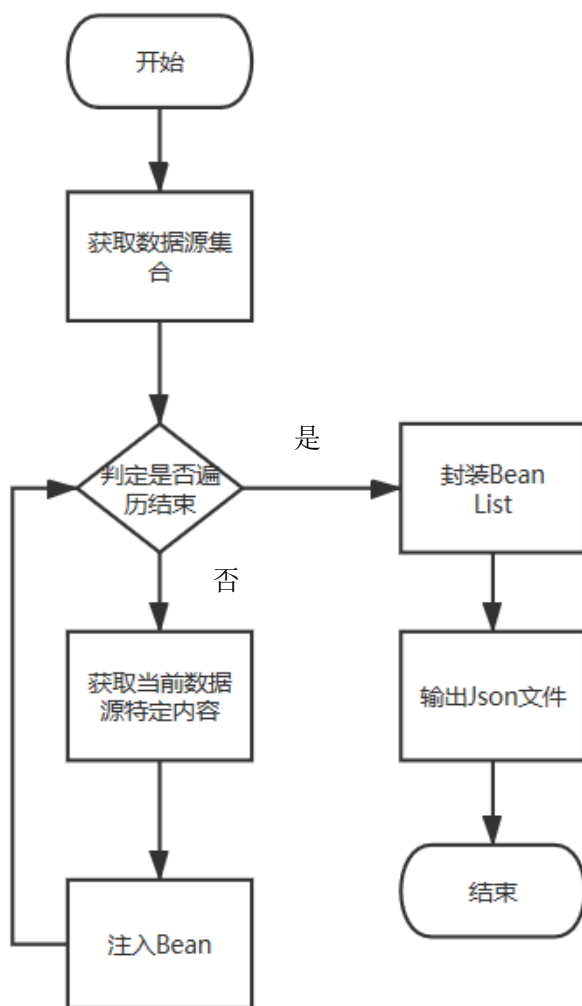


图4.6 结构化数据抽取流程图

非结构化数据抽取方面，使用正则表达式进行模式匹配，对此前预处理后的数据文件进行特征提取，并将各条目信息注入 java bean 中，对 bean 进行组织，最后使用序列化方法对其进行格式化。算法流程图如图 4.7 所示。

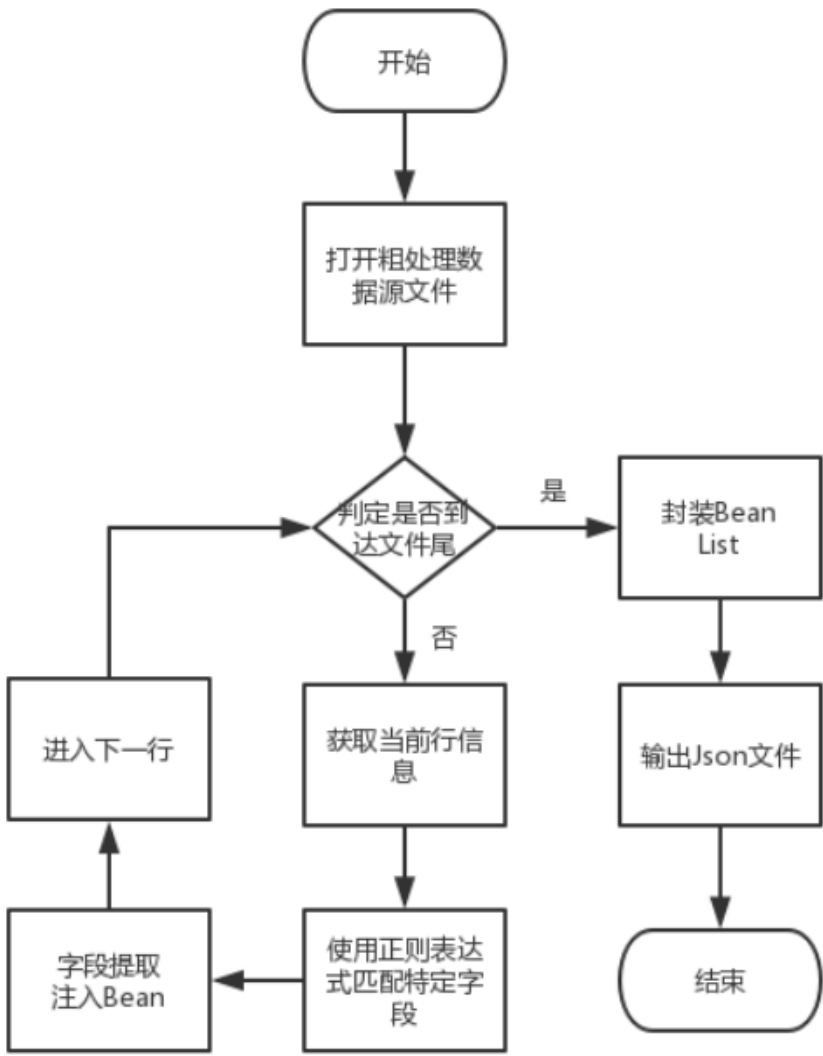


图4.7 非结构化数据抽取流程图

5. 设计成果

5.1 系统主界面

本系统基于 ASP 技术开发，是 B/S 架构的 Web 应用程序。在前端部分我使用了 Bootstrap 框架实现响应式页面显示，并嵌入子页面实现在单一窗口进行操作。

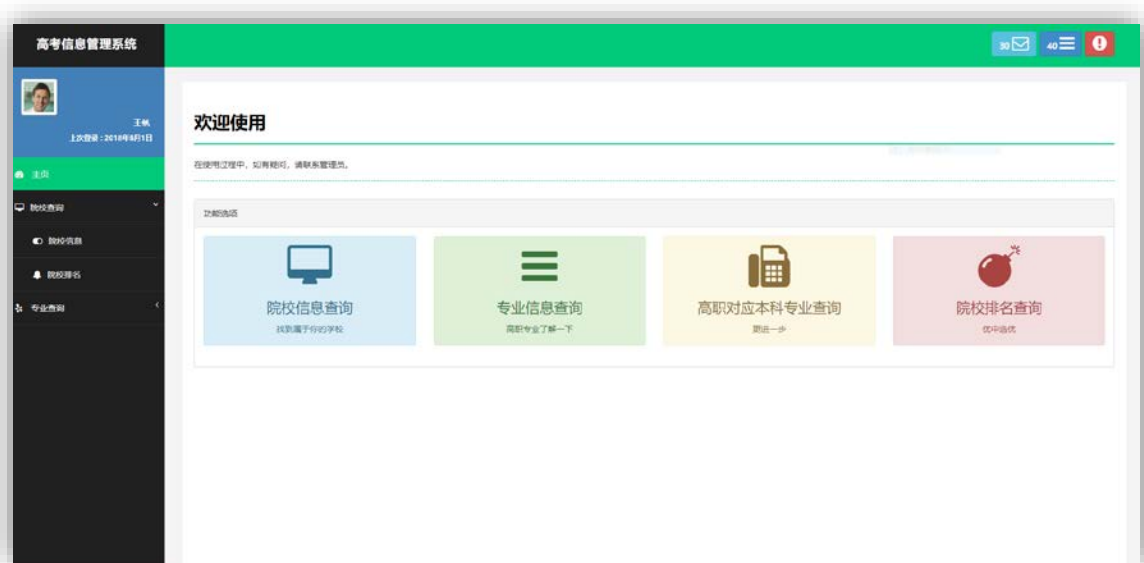


图5.1 系统主界面（1920×1080分辨率）

主页面提供了较为友好的人机交互提示，并在左侧边栏进行栏目切分，便于用户选择对应的选项。

5.2 院校信息检索

院校信息是本系统中较为关键的信息实体。本系统通过对基础数据表进行视图化处理，实现使用较为简捷的查询方法获取信息的操作。

此外，针对用户不同的需求，对信息检索的选项也需要加以考量。本系统提供了院校名称模糊检索、所在地检索、学校类型筛选检索等方式，可以满足用户的不同需求，并提供了较好的扩展性。



图5.2.1 院校查询主界面

院校代码	院校名称	学科类别	所在省份
4111014279	北京交通职业技术学院	综合类	北京
4112012484	天津滨海职业学院	综合类	天津
4112012719	天津渤海职业技术学院	综合类	天津
4112012722	天津现代职业技术学院	综合类	天津
4112012732	天津轻工职业技术学院	综合类	天津
4112012803	天津国土资源职业学院	综合类	天津
4112013702	天津城市职业学院	综合类	天津
4112014020	天津城市建设管理职业技术学院	综合类	天津
4112014022	天津海运职业学院	综合类	天津
4113011238	石家庄职业技术学院	综合类	河北
4113012415	沧州职业技术学院	综合类	河北
4113012543	保定职业技术学院	综合类	河北
4113012667	河北旅游职业学院	综合类	河北

图5.2.2 院校查询结果列表

选择对应院校，可以跳转到院校详情页，如下图所示。

北京交通职业技术学院介绍	
院校名称:	北京交通职业技术学院
院校类型:	公办
院校代码:	4111014279
学科类别:	综合类
主管部门:	北京市教育委员会
所在省份:	北京
全国排名:	465
省内排名:	12
评分等级:	3★
简介:	学院简介日期: 2017-04-27 09:49 39北京交通职业技术学院是北京市交通系统属的唯一公办全日制普通高等职业院校。市交通委主任招
收费标准:	4. 学费标准: 6000元/年/人, 住宿费标准: 大一校区750元/年/人, 大二校区650元/年/人, 海岱校区650元/年/人, 1200元/年/人。
官方网站:	http://www.bjtt.edu.cn
曾用名:	北京交通职业技术学院
开设专业:	安全技术与管理 古建筑工程技术 建设工程监理 汽车检测与维修技术 汽车电子技术 新能源汽车技术 智能交通技术运用 道路桥梁工程技术 道路运输与路政管理 交通规划运营管理 汽车运用与维修技术 汽车车身维修技术 民航运输 航空物流 城市轨道交通车辆技术 城市轨道交通机电技术 城市轨道交通通信信号技术 城市轨道交通供电技术 城市轨道交通工程技术 城市轨道交通运营管理 汽车营销与服务 物流管理

图5.2.3 院校详细信息

选择对应开设专业，可以跳转到专业详情页，如下图所示。



图5.2.4 专业详细信息

检索页主要通过使用 TextBox 等前端控件实现对数据的装载，并通过 From 表单进行数据传递。使用 ASP 技术在查询结果页进行数据的封装处理，获取数据库信息，并进行格式化输出操作。

在每行加入跳转至相应院校详情页面的链接，在院校详情页开设专业项中，为每个专业提供超链接跳转。

5.3 专业信息检索

专业信息是本系统中较为关键的信息实体。本系统通过对基础数据表进行视图化处理，实现使用较为简捷的查询方法获取信息的操作。

此外，针对用户不同的需求，对信息检索的选项也需要加以考量。本系统提供了专业名称模糊检索、专业代码检索等方式，可以满足用户的不同需求，并提供了较好的扩展性。

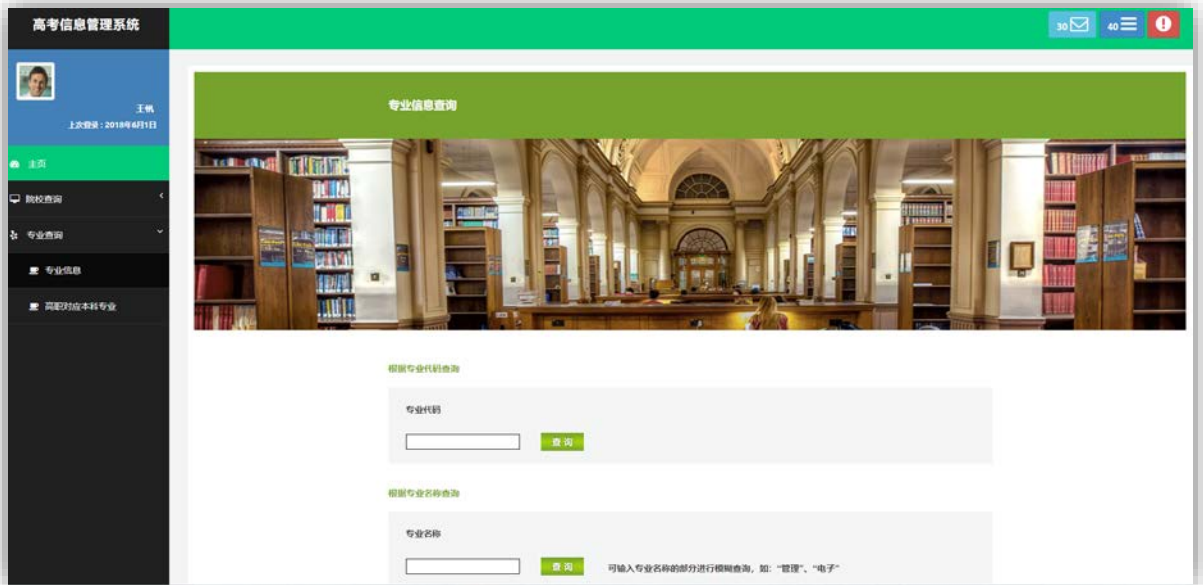


图5.3.1 专业查询主界面



图5.3.2 模糊查询结果



图5.3.3 专业详细信息

检索页主要通过使用 `TextBox` 等前端控件实现对数据的装载，并通过 `Form` 表单进行数据传递。使用 `ASP` 技术在模糊查询结果页进行数据的封装处理，获取数据库信息，并进行格式化输出操作。

在专业详情页对应本科专业一项中，使用视图技术获取对应本科专业信息，并格式化输出。

5.4 院校排名信息

院校排名信息是本系统中较为关键的信息实体。本系统通过对基础数据表进行视图化处理，实现使用较为简捷的查询方法获取信息的操作。由于院校排名时效性高，因此使用视图的结构对元数据进行透视操作，以便日后更新排名信息。

本模块使用对院校排名视图的单表查询并在每行中加入超链接跳转，然后结构化显示。



图5.4 院校排名信息

3.5 高职院校对应本科专业信息

高职院校对应本科专业信息是本系统较为特色的信息。由于高职高专考生存在日后专升本的潜在可能性，而目前市面上同类型应用并未提供相关的内容，因此通过本系统可以对特定高职专业与本科专业相对应信息进行查询，从而方便地获取到专升本信息。

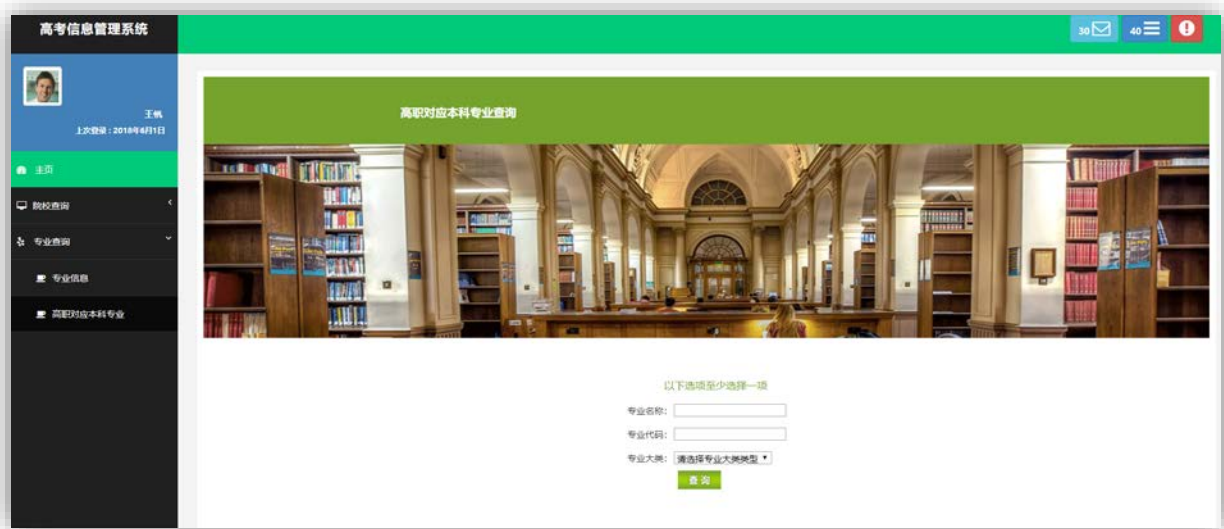


图5.5.1 高职对应本科专业信息查询主界面



图5.5.2 高职对应本科专业信息查询结果

检索页主要通过使用 TextBox 等前端控件实现对数据的装载，并通过 From 表单进行数据传递。使用 ASP 技术在模糊查询结果页进行数据的封装处理，获取数据库信息，并进行格式化输出操作。

在查询结果页中对每行加入超链接，方便用户获取专业详情信息。

6. 总结

在实现本系统的过程中，通过查找大量的资料和通过对其他作品的学习，对 ASP 的使用有了基础的认识，能够在短时间内写出一个网站的基本框架。在解决数据存储的过程中，在 MariaDB 数据库的基础上，了解并学习了 Navicat 等数据库管理软件的使用方法，并实现了所需功能。

在开发过程中，我学习了爬虫相关的知识，从中国教育在线^[6]与全国职业院校专业设置管理与公共信息服务平台获取到了本课程设计所需要的相关原始数据，并进行了整理，以便使用。

在开发过程中，我也发现了 ASP 的一些不足之处。虽然 ASP 可以在短时间内较简单的开发出所需的功能，但由于历史原因，其扩展性较差，且没有主流的 IDE 辅助开发。不过作为一种快速开发语言，ASP 有着较为简捷与便利的优点，在本项目中也有着较为良好的表现。

为了实现较人性化的前端界面，我学习使用了目前流行的前端库 Bootstrap v3，运用该框架使得本系统可以实现响应式网站设计，以及其他较为美观的效果。

本专业设计只是完成了高校信息查询系统的基本必备功能：对院校、专业数据的条件检索与显示功能，但还不够完善。在以后的学习中，应完成的工作是对这些功能的细节补充以及架构重构，例如数据源更新、用户鉴权、部分时效性数据定期自动维护等。

7. 参考文献

- [1] 曹亚辉. 非结构文本最佳近似匹配系统的研究与实现：[硕士学位论文]. 东华大学, 2011.
- [2] OASIS 成员通过非结构化信息访问开放标准. <http://www.oasis-open.org/cn/news/2009-03-19.htm>. 2010-07-15
- [3] 邱明辉, 彭强. 非结构化信息管理框架的原理与应用. 情报探索, 2011, 01: 93-95
- [4] 霍焰. 基于非结构化文档数据的抽取与分析系统的信息抽取：[博士学位论文]. 天津大学, 2012.
- [5] Approximate string matching Algorithm. Narendra Kumar, Vimal Bibhu, Mohammad Islam, Shashank Bhardwaj. International Journal on Computer Science and Engineering . 2010
- [6] 詹志建, 杨小平. 基于语言网络和语义信息的文本相似度计算. 计算机工程与应用, 2014, 50(05): 33-38.

二 专业设计过程记录表

时间	完成的内容	指导教师意见
2016 年 11 月	了解并学习 Java 程序设计语言 使用 Java 开发数据集成系统。	
2017 年 5 月	了解并学习数据库相关知识，实 现嵌入式数据库应用程序开发。	
2017 年 11 月	了解并学习爬虫相关知识，使用 Jsoup 实现数据源数据抽取工作， 研究正则表达式在匹配中的使 用方法。	
2018 年 5 月	系统学习 HTML/CSS/JavaScript 相关知识，实现前端页面搭建。	
2018 年 11 月	学习 ASP，实现动态交互 B/S 系 统搭建，完成专业设计报告。	

教师签字：