

高考志愿查询预测与推荐系统

数据库设计文档

项目名称：高考志愿查询预测与推荐系统

项目组：高考志愿查询预测与推荐系统开发组

项目成员：陈思翰（项目经理）、李响、薛安宇、胡周峰、田如召

2021 年 6 月 2 日

目 录

1. 引言.....	1
1.1 文档编写目的.....	1
1.2 文档编写背景.....	1
1.3 文档适用范围.....	1
1.4 相关定义.....	1
2. 外部设计.....	1
2.1 表的划分.....	1
2.2 标识符	2
2.3 数据库表设计格式.....	2
2.3.1 高校总表.....	2
2.3.2 高校分数线总表.....	2
2.3.3 专业分数线总表	2
2.3.4 专业总表.....	2
2.3.5 专业明细表	3
2.4 支持软件.....	3
3. 结构设计.....	3
3.1 概念结构设计.....	3
3.2 逻辑结构设计.....	3
4. 运用设计.....	4
4.1 数据字典设计.....	4
4.1.1 数据项.....	4
4.1.2 数据结构.....	4
4.1.3 外部实体.....	4
4.2 安全保密设计.....	4

文档信息	
标题	数据库设计文档 Database Design Document
作者	薛安宇
创建日期	2021.05.24
更新日期	2021.06.02
版本	V1.3
部门名称	高考志愿查询预测与推荐系统项目开发组-华中农业大学实训第五组

文档更新记录			
版本号	更新内容	操作日期	操作人员
1.0	初步拟定文档，并填入相关信息	2021.05.24	薛安宇
1.1	补充了 2.2 节标识符和第 4 部分运用设计	2021.05.25	田如召
1.2	更新文档框架，使其视觉效果更加美观	2021.05.26	薛安宇、陈思翰
1.3	更新了 4.1.2 节 数据结构	2021.06.02	陈思翰

数据库设计文档

Database Design Document

1. 引言

1.1 文档编写目的

本《数据库设计文档》旨在确立本项目数据所需的存储方式和具体数据库以及数据库中需要建立的表和字段，以便后续与前端部分进行高效对接。

1.2 文档编写背景

针对“高考志愿查询预测与推荐系统”项目，需要建立数据库以存储开发过程中需要用到的上百万条数据，因此编写此文档。

1.3 文档适用范围

本文档主要针对进行数据库建立与维护以及相关接口开发的项目组成员。其他项目组成员也可以阅读该文档，便于更深入的了解该项目。

1.4 相关定义

①数据库：数据库是“按照数据结构来组织、存储和管理数据的仓库”，是一个长期存储在计算机内的、有组织的、可共享的、统一管理的大量数据的集合。

②表：表是数据库中用来存储数据的对象，是有结构的数据的集合，是整个数据库系统的基础。

③字段：一个成员，它表示与对象或类关联的变量。在数据库中，大多数时，表的“列”通常被称为“字段”，每个字段包含某一专题的信息。

2. 外部设计

2.1 表的划分

- ①高校总表：记录所有高校的相关信息；
- ②高校分数线总表：记录所有高校在每个地区的历年分数线；
- ③专业分数线总表：记录各高校各专业在每个地区的历年分数线；
- ④专业总表：记录全国高校开设的所有专业的简略信息；
- ⑤专业明细表：记录所有专业的详细信息。

2.2 标识符

- ①高校总表: SchoolTable
- ②高校分数线总表: FractionalTable
- ③专业分数线总表: ScoreInformation
- ④专业总表: ProfessionalTable
- ⑤专业明细表: SubsidiaryTable

注: 本项目所有表均无标识符, 且后续开发过程中可能会进行变更。

2.3 数据库表设计格式

2.3.1 高校总表

编码: utf-8

字段名	类型	字段名	类型
高校名称	varchar	招生办网址	varchar
高校层级	varchar	学校简介	varchar
高校类别	varchar	学生评论 1	varchar
高校所在地	varchar	学生评论 2	varchar
占地面积	int	学生评论 3	varchar
建校年份	int	学生评论 4	varchar
电子邮箱	varchar	学生评论 5	varchar
联系电话	varchar		

2.3.2 高校分数线总表

编码: utf-8

字段名	类型	字段名	类型
高校名称	varchar	招生类型	varchar
省份	varchar	最低分	int
科类	varchar	最低位次	int
年份	int	省控线	int
录取批次	varchar	选课要求	varchar

2.3.3 专业分数线总表

编码: utf-8

字段名	类型	字段名	类型
高校名称	varchar	招生类型	varchar
省份	varchar	最低分	int
科类	varchar	最低位次	int
年份	int	省控线	int
录取批次	varchar	选课要求	varchar
专业名称	varchar		

2.3.4 专业总表

编码: utf-8

字段名	类型	字段名	类型
专业名称	varchar	专业门类	varchar
专业类别	varchar	专业大类	varchar

2.3.5 专业明细表

编码: utf-8

字段名	类型	字段名	类型
专业门类	varchar	考研方向	varchar
专业大类	varchar	发展前景	varchar
专业层次	varchar	2015 年就业率	varchar
专业简介	varchar	2016 年就业率	varchar
主修课程	varchar	2017 年就业率	varchar
发展方向	varchar	就业行业分布	varchar
修读年限	varchar	就业地区分布	varchar
男女比例	varchar	就业岗位分布	varchar
文理比例	varchar		

2.4 支持软件

- ①操作系统: Windows 操作系统、Mac OS;
- ②数据库: MongoDB;
- ③编程工具: PyCharm;
- ④数据处理: Spark;
- ⑤MongoDB 可视化工具 Robo 3T。

3. 结构设计

3.1 概念结构设计

- ①管理员——更新数据
- ②用户——浏览数据

3.2 逻辑结构设计

①高校数据: 高校名称、高校层级、高校类别、高校所在地、占地面积、建校年份、电子邮箱、联系电话、招生办网址、学校简介、学生评论 1、学生评论 2、学生评论 3、学生评论 4、学生评论 5;

②高校分数线数据: 高校名称、省份、科类、年份、录取批次、招生类型、最低分、最低位次、省控线、选课要求;

③专业分数线数据: 高校名称、省份、科类、年份、录取批次、专业名称、招生类型、最低分、最低位次、省控线、选课要求;

④专业数据: 专业名称、专业类别、专业门类、专业大类;

⑤专业明细数据: 专业门类、专业大类、专业层次、专业简介、主修课程、发展方向、修读年限、男女比例、文理比例、考研方向、发展前景、2015 年就业率、2016 年就业率、2017 年就业率、就业行业分布、就业地区分布、就业岗位分布。

4. 运用设计

4.1 数据字典设计

4.1.1 数据项

与 2.3.1~2.3.5 节一致。

4.1.2 数据结构

与 3.2 节一致，其中：

- ①高校数据表：记录了全国高校的各种简略信息；
- ②高校分数线表：记录了全国各高校在各地区历年来的最低分数线和最低排名；
- ③专业分数线表：记录了全国各高校开设的各专业在各地区历年来的最低分数线和最低排名；
- ④专业数据表：记录了全国所有专业的各种简略信息，用于与表⑤进行比对填补确实数据；
- ⑤专业明细表：记录了全国所有专业的详细信息；
- ⑥高校分数线预测表：记录了全国所有高校在各地区的 2021 年预测分数线。

4.1.3 外部实体

名称：数据库工程师/管理员

说明：管理数据库后台的主要实体

输入数据：新爬取的数据流

输出数据：清洗后的数据项

4.2 安全保密设计

暂无。