

具有推荐功能的志愿填报助手的设计与实现

目录

1

意义与需求

系统设计

3

系统实现

结果与总结

2

4

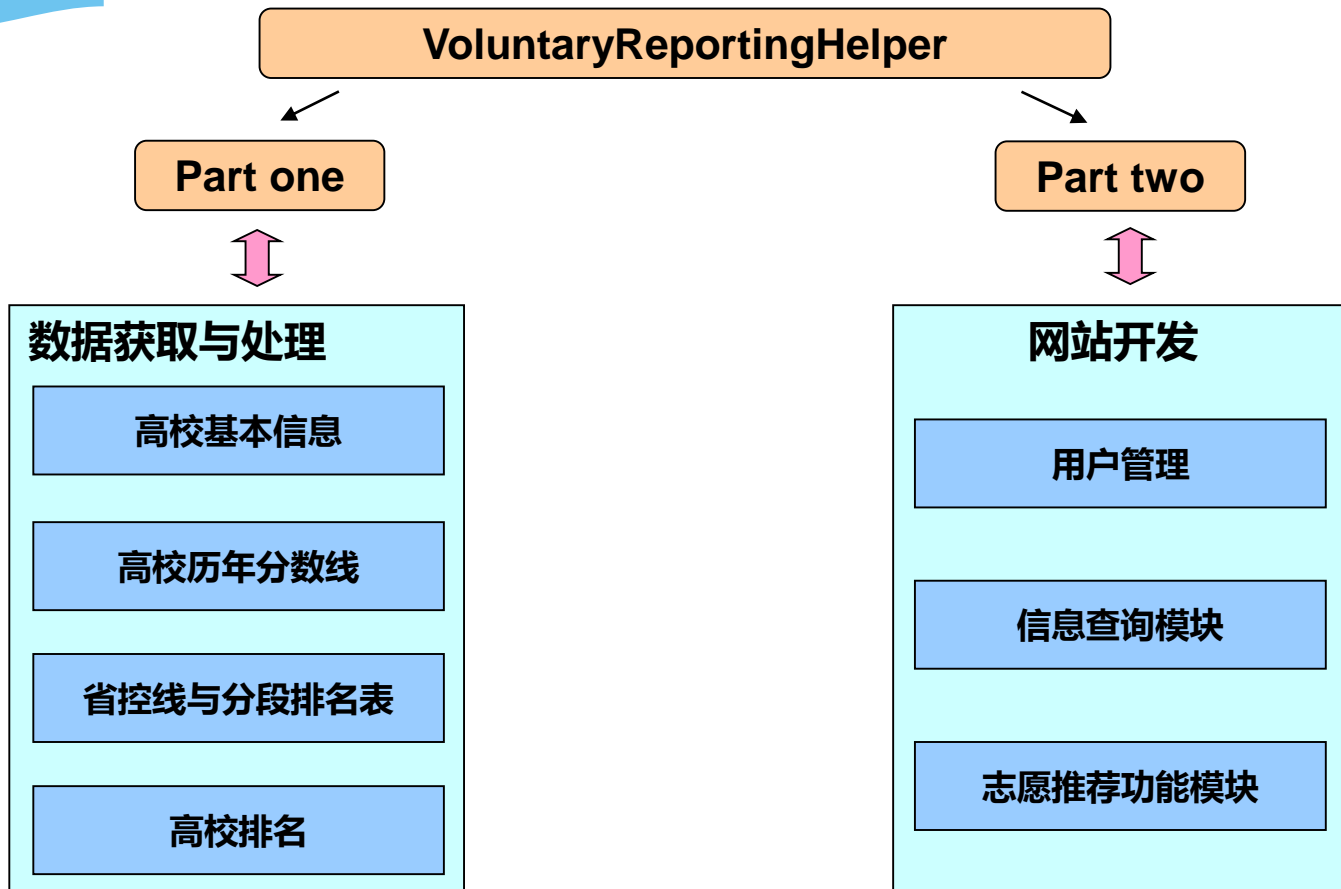
意义

1. 高考志愿填报至关重要，是一场博弈。
2. 有效的信息搜集，依据分数做决策。
3. 众多影响因素，城市、分数增长、属性、就业等。
4. 基于用户喜好，考虑多方面，遵循择校理念。
5. 个性化的志愿推荐，极具意义。
6. 对考生和家长的志愿填报有重要的指导作用。

需求

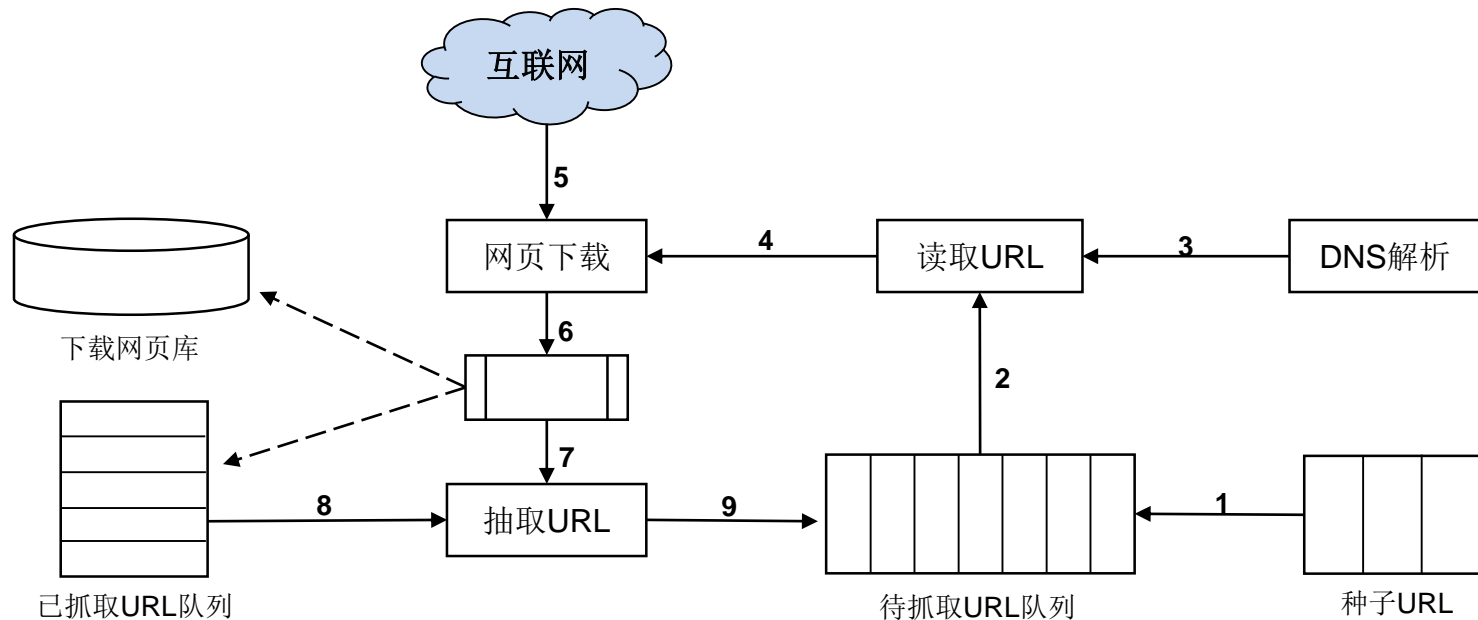
1. **数据**：各省批次线，高校历年录取分数，省分段排名表，学校基本信息（属性，类型，城市，学科建设，师资，专业设置，就业，保研等等）。——设计爬虫，从各大网站爬取。真实！
2. **查询**：高校信息，各类排名，各种分数，志愿百科文章。
3. **志愿推荐**：基于分数，对择校影响因素指标的关心程度建立“指标权重模型”，模型下的加权得分——匹配度得分 → 推荐学校 → 推荐填报方案

开发流程



数据获取

1. 分数信息：----中国教育在线网，51EDU网，高考志愿网
2. 高校基本信息：----高校百度百科，高考派
3. 各类排名信息：----青塔，新浪高考
4. 爬虫设计：python2.7+urllib2+selenium+正则表达式



数据存储

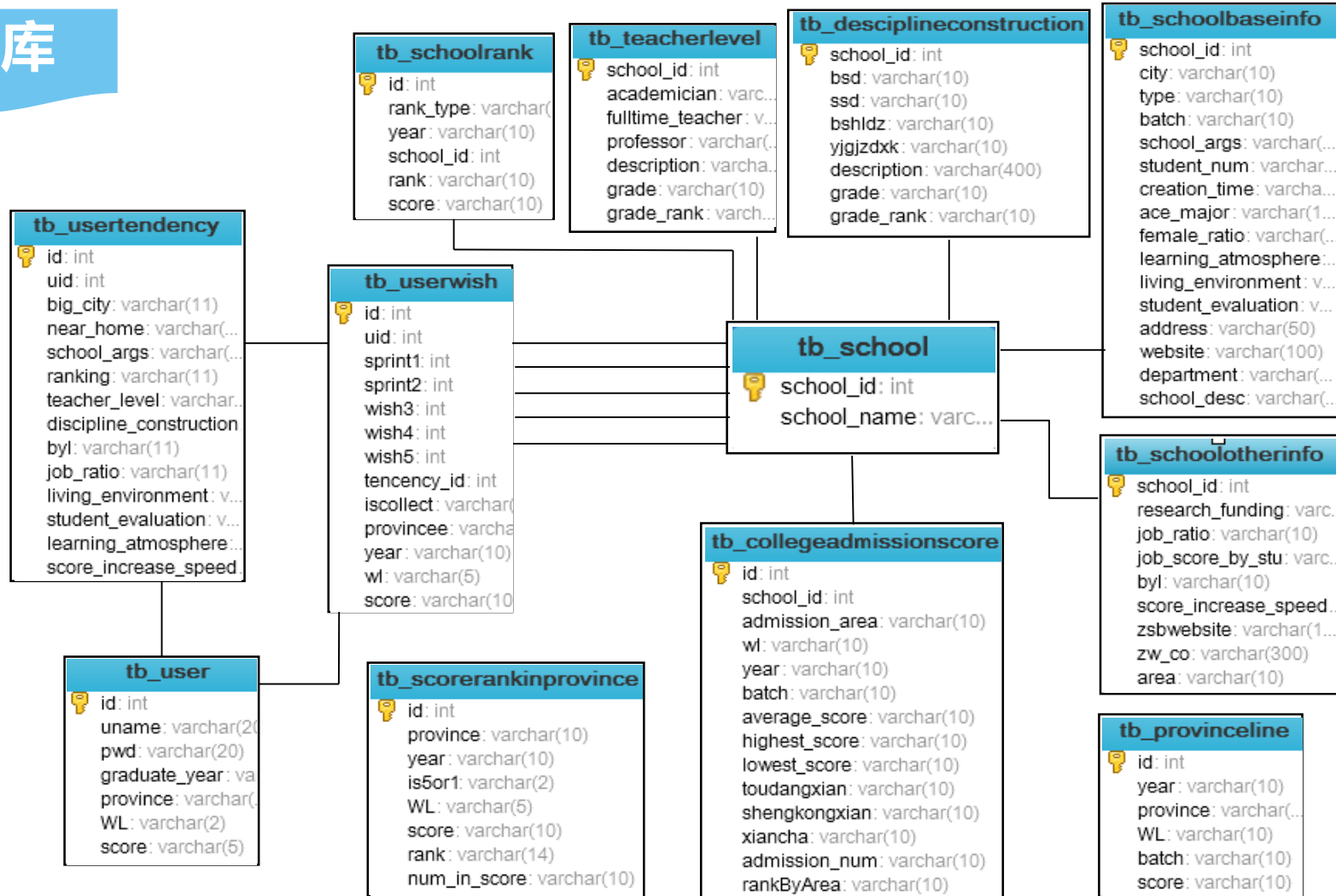
JSON存储，数据有120万+。示例：

```
{  
  "schoolid": "31",  
  "schoolname": "北京大学",  
  "localprovince": "江苏",  
  "studenttype": "文科",  
  "year": "2015",  
  "batch": "本科一批",  
  "var": "405",  
  "var_score": "405",  
  "max": "416",  
  "min": "398",  
  "num": "0",  
  "fenchu": "63",  
  "provincescore": "342"  
}
```

电脑 > 本地磁盘 (D:) > Python27 > Mydata > MyBigData > Zzz

名称	修改日期	类型	大小
RANKs	2016/5/30 23:24	文件夹	
AllschoolcitiaoURLs.json	2016/5/24 10:53	JSON 文件	270 KB
Batch1SchoolInfo_Old.json	2016/5/24 11:10	JSON 文件	676 KB
gaokaopaischoolNos.json	2016/5/24 10:36	JSON 文件	15 KB
GKPcollegeDesc.json	2016/5/30 1:42	JSON 文件	2,675 KB
gkpSchoolBaseInfo.json	2016/5/24 14:28	JSON 文件	222 KB
GKPshengkongxian.json	2016/5/23 1:21	JSON 文件	115 KB
GKZY100_all	2016/6/1 18:25	文件	3,770 KB
GKZY100_all_batch2	2016/6/7 23:42	文件	9,687 KB
GKZY100_all_batch3	2016/6/8 0:14	文件	3,983 KB
GKZY100_all_batch4	2016/6/8 1:21	文件	783 KB
gaoxiao51eduInfo.json	2016/5/7 18:09	JSON 文件	13,805 KB
_20151fenduan.json	2016/5/25 13:33	JSON 文件	141 KB
collegeInfo.json	2016/5/18 0:20	JSON 文件	30,727 KB
collegeInfo_score.json	2016/5/10 11:28	JSON 文件	30,973 KB

数据库



算法设计

决策影响因素：

1. 所在城市
2. 离家远近
3. 排名
4. 属性 (985,211)
5. 就业
6. 师资力量
7. 学科建设
8. 学生评价
9. 学习氛围
10. 生活环境
11. 分数增长率
12. 保研率

数据挖掘

量化

各高校
各项指标
的评分

- | | | |
|---------|---|------|
| 1. 所在城市 | → | 发达程度 |
| 2. 离家远近 | → | 省间距离 |
| 3. 学生评价 | → | 综合多值 |
| 4. 分数增长 | → | 挖掘信息 |

.....

算法设计

志愿推荐：

1. 高考分数 \longrightarrow 可能录取的学校集合S

2. 选择对各指标的关心程度 $\xrightarrow[\text{1 / 0.5 / 0}]{\text{很 / 一般 / 不}}$ “指标权重模型” M

3. 计算S中所有学校在M下的加权得分 $\xrightarrow{\text{相加}}$ 匹配度得分（相似性）

4. S的所有学校匹配度得分 $\xrightarrow[\text{理由分析}]{\text{排序}}$ 推荐的高校列表L

5. L中根据2冲2稳1保 $\xrightarrow{\text{风险规避}}$ 推荐的志愿填报方案

功能图

VoluntaryReportingHelper

part1

part2

part3

信息查询模块

高校基本信息

高校录取分数线

各省历年批次线

各省历年分段表

各类排名

志愿百科文章

志愿推荐模块

志愿填报方案推荐

志愿方案收藏

志愿方案查看

用户模块

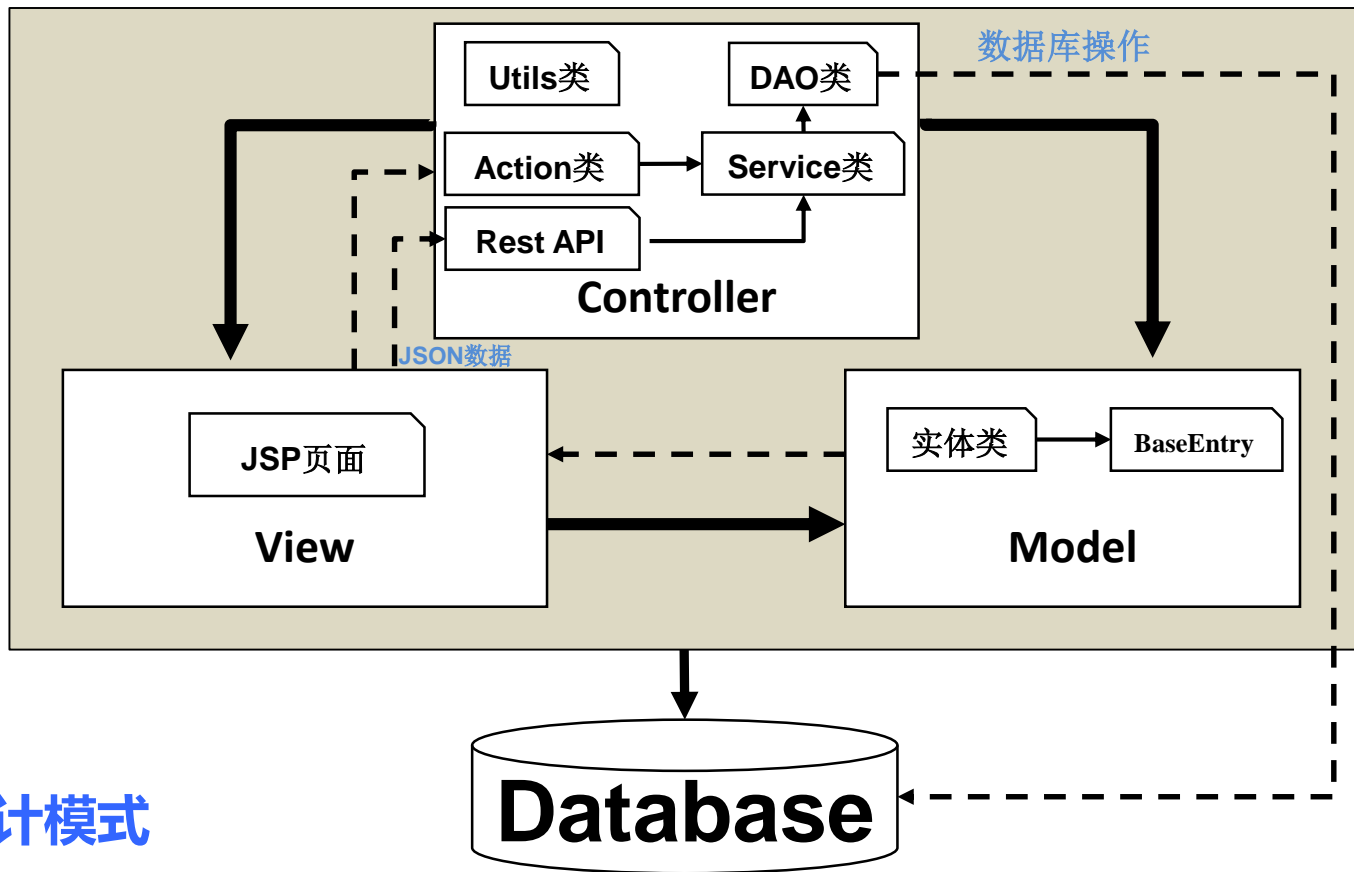
用户注册

用户登录

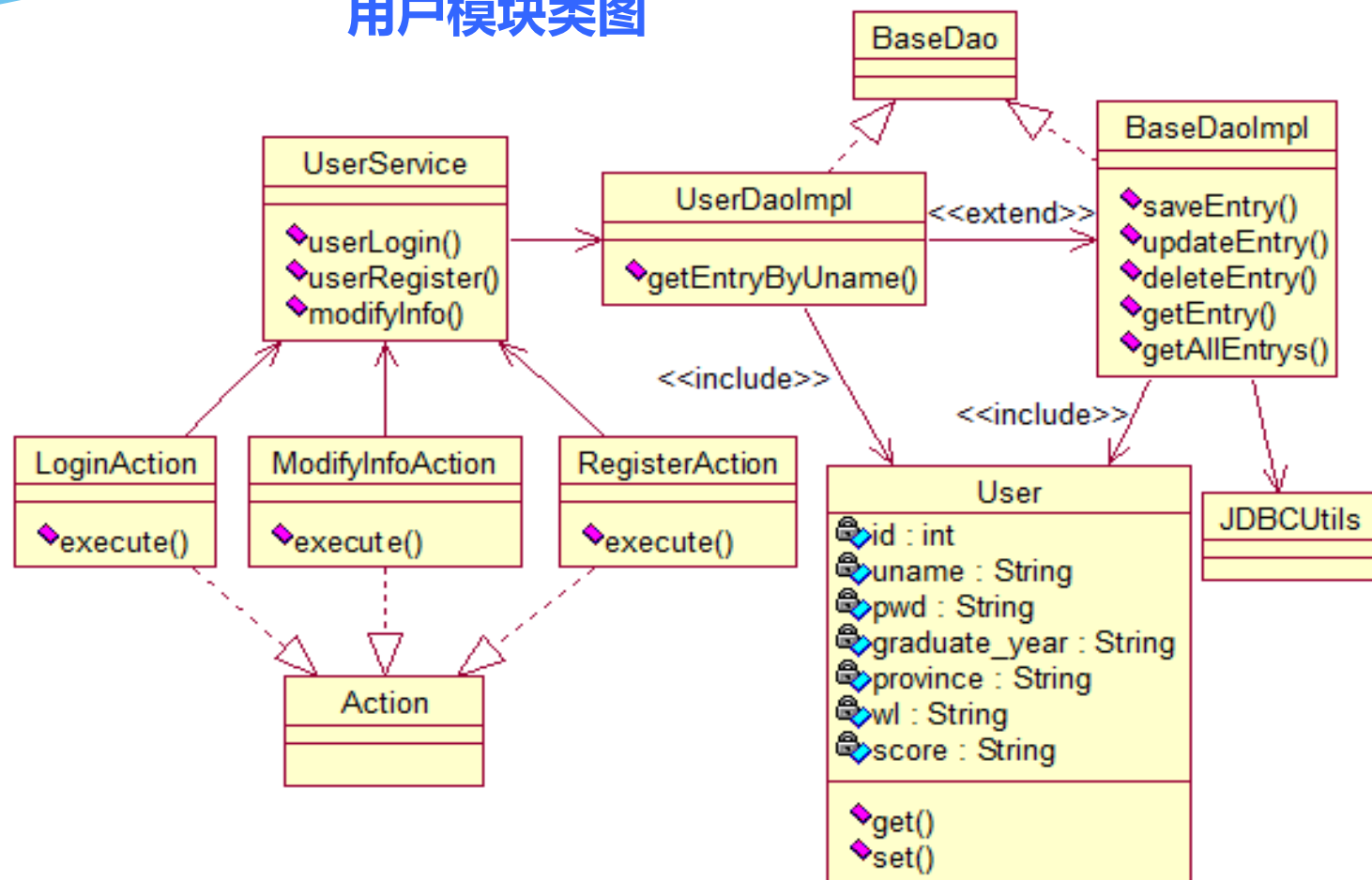
用户注销

用户信息修改

VoluntaryReportingHelperDB



用户模块类图



注册信息

*账号:

*密码:

*毕业年份:

*毕业省份:

*文理科: ☐ 文 ☒ 理

*分数:

注册

注: 注册成功后资料不可更改。

localhost:8090/WishHelper/



将来的你
一定会感谢现在拼命的自己

用户: 未登录

搜索



高考志愿填报助手登录

账号:

密码:

验证码:  [看不清?换一张](#)

登陆

重置

[没有账号?](#) | [忘记密码?](#)

展示

主页



用户 : sa

搜索 →

大学库

高校录取线

各省批次线

高考分数段排名

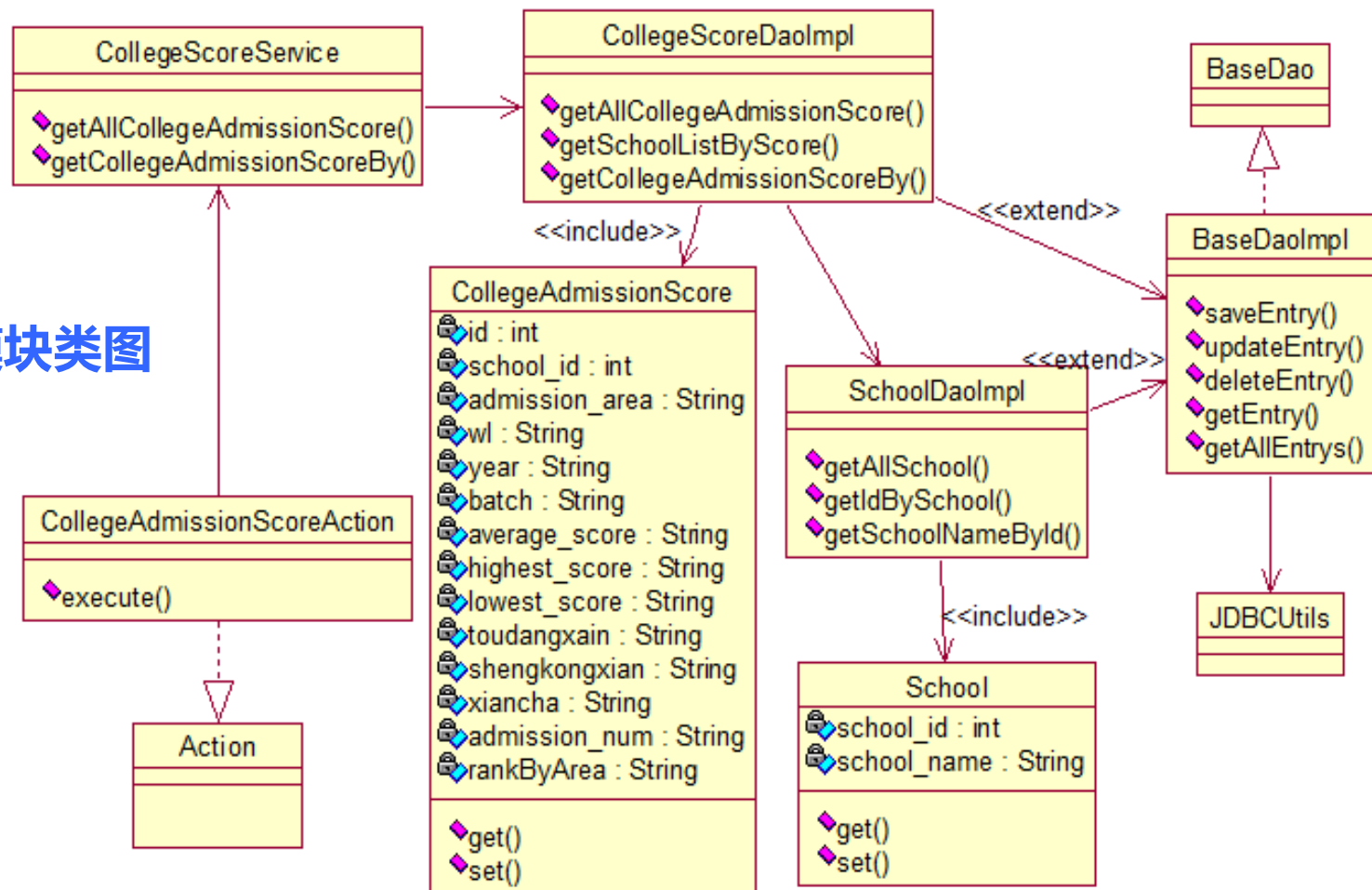
志愿推荐

大学排名

志愿百科



查询模块类图



2015年各高校在北京录取线

学校名称:

招生省份:

北京 ▼

文理科:

理科 ▼

录取批次:

本科一批 ▼

年份:

2015 ▼

查询

院校名称	招生地区	文理科	招生年份	招生批次	平均分	最高分	最低分	省控线	录取人数
上海交通大学	北京	理科	2015	本科一批	688	--	--	548	0
浙江大学	北京	理科	2015	本科一批	687	--	--	548	0
中国人民大学	北京	理科	2015	本科一批	686	--	--	548	0
北京师范大学	北京	理科	2015	本科一批	680	--	--	548	0
中央财经大学	北京	理科	2015	本科一批	674	--	--	548	0
厦门大学	北京	理科	2015	本科一批	670	--	--	548	0
上海财经大学	北京	理科	2015	本科一批	670	--	--	548	0
中山大学	北京	理科	2015	本科一批	668	--	--	548	0
北京理工大学	北京	理科	2015	本科一批	667	--	--	548	0



学校简介

哈尔滨工业大学创建于1920年,坐落在素有“东方莫斯科”之称的北国名城哈尔滨市。隶属于国防科工委。经过87年的建设,现已发展成为一所理工为主,理、工、管、文、经、法相结合,在国际上享有一定声誉的开放式、研究型全国著名重点大学。

在各个历史时期,哈工大的发展始终受到国家的重点支持。20世纪50年代,哈工大是我国政府确定的学习苏联先进教育制度的两所院校之一。自1954年成为国家确定的首批6所重点大学之一以来,一直是国家重点大学。1984年又被国家列为重点建设的15所大学之一,1996年成为首批进入国家“211工程”重点建设的院校之一,1999年被确定为国家“985工程”重点建设的9所大学之一。2000年6月,哈尔滨工业大学与同根同源的哈尔滨建筑大学合并,组成新的哈尔滨工业大学,两校合

基本信息

创建时间	1920年	学校类型	理工类
主管部门	工业和信息化部	学校性质	985 211
所在城市	黑龙江哈尔滨	学校地址	哈尔滨市南岗区西大直街92号
学生人数	3万~4万人	学校官网	http://www.hit.edu.cn/
校园面积	未知	男女比	--

王牌专业

飞行器设计与工程,探测制导与控制技术,焊接技术与工程,轮机工程,飞行器动力工程,数学与应用数学,工商管理,建筑环境与能源应用工程,道路桥梁与渡河工程,化学工程与工艺

师资力量

截止2015年,校本部有专任教师2970人,其中教授930人;有长江学者58人,国家杰青39人,国家级教学名师9人,国家级教学团队6个,教育部创新团队12个。[17]

专职教师	2970人	教授	930人
两院院士	33人	师资评分:	8.0分/排名:24

学科建设

截止2015年,学校有一级学科国家重点学科9个,二级学科6个,二级国家重点(培育)学科2个,有一级学科博士学位授予点27个,博士专业学位授权点2个;一级学科硕士学位授予点41个,硕士专业学位授权点30个;博士后科研流动站25个。

博士点	82个	博士后流动站	25个
硕士点	148个	国家重点学科	40个
学科评分:	8.07分/排名:24		

2016中国理工类大学排行榜

排名类型 中国大学理科录取分数2005-2014十年总排名 ▼

年份: 2016 ▼

查询

排名	学校	年份	得分
1	清华大学	2016	98.50
2	国防科学技术大学	2016	80.31
3	中国科学技术大学	2016	75.14
4	华中科技大学	2016	75.12
5	同济大学	2016	72.85
6	天津大学	2016	72.81
7	哈尔滨工业大学	2016	72.72
8	北京航空航天大学	2016	70.58
9	东北大学	2016	69.55
10	大连理工大学	2016	68.84
11	北京理工大学	2016	68.72

填报技巧

- 1.志愿填报中的陷阱——专业级差
- 2.填报志愿，更应该关注专业选择
- 3.志愿“家长包报”弊端多

专业选择

- 1.如何选专业，关注专业培养方向，注意大类招生
- 2.倾情奉献！非211大学中的好大学好专业！
- 3.知名校长谈如何选专业

大学选择

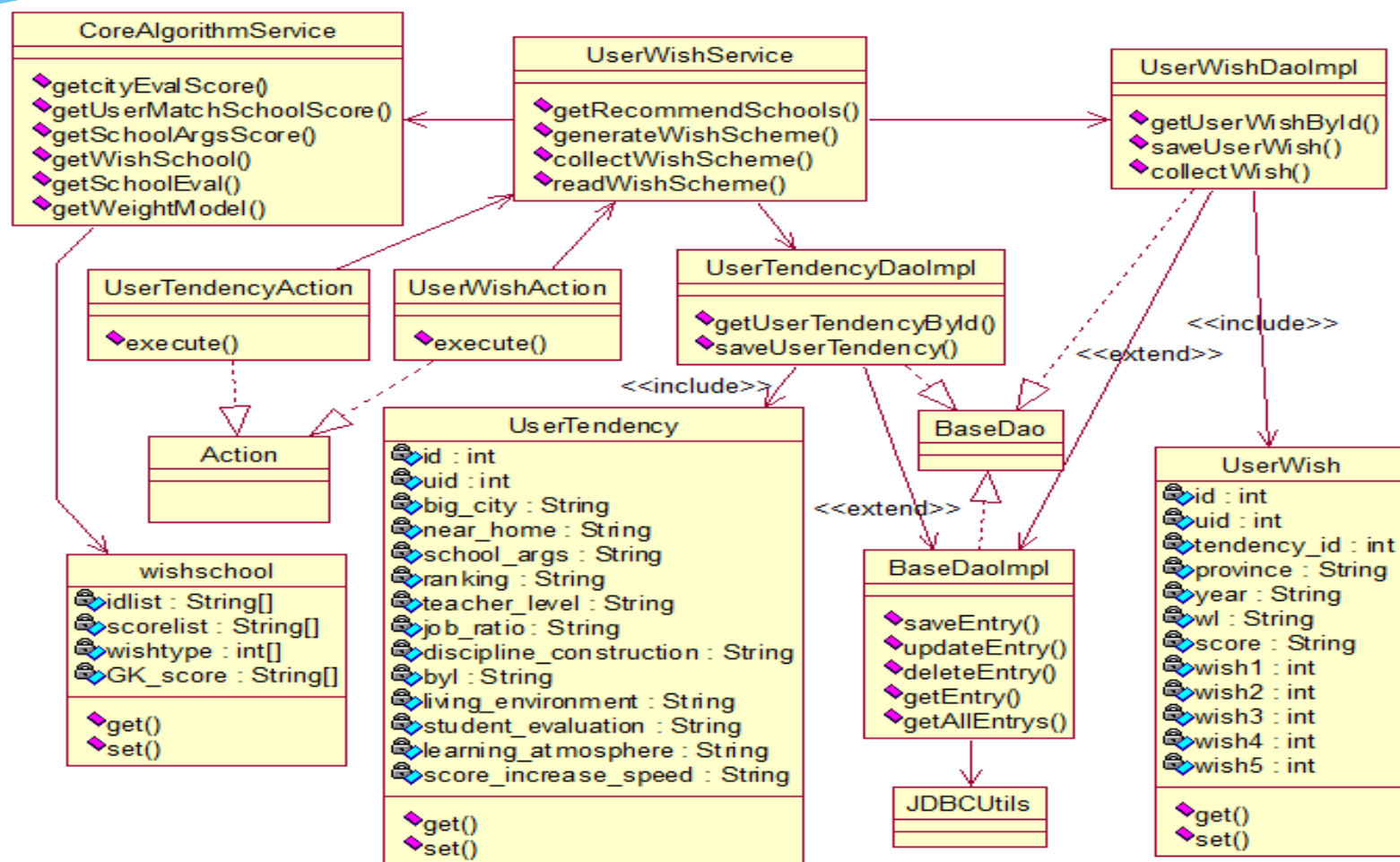
- 1.中国各行业最牛的十所大学
- 2.盘点985和211中的“工业”大学
- 3.盘点十大理科高考最难考的院校

在整个高考志愿填报的过程中,高校常见的三种专业录取规则分别为:分数优先、专业优先(专业清)以及专业级差三种方式。对于前两者,大家或许并不陌生,但专业级差可能就没听过了,那什么是专业级差呢?专业级差是指在第一专业没有被录取,再按第二专业录取时要减去一个分数。如考生王某高考分550分,其专业志愿分别是工商管理、电子工程、财务管理、信息工程、国际经济与贸易、保密管理等六个专业,当第一志愿专业“工商管理”不能满足,而所报院校的专业级差为3分,该生就要按547分的规定成绩同第一志愿报考“电子工程”的考生的高考原始分排队再安排专业。



陷阱(1):专业报考时留意设立专业级差的院校。高校设置的专业级差,主要目的是让考生在选择志愿时要冷热搭配,全面考量,综合分析,理智选择。如果一门心思专选热门专业,分数达不到还减去一定的分数,可能会录取到最冷门的专业。如上海财经大学采取33111的级差进行专业志愿的录取,就是说第二、三专业志愿的考生每一档降3分与前一专业志愿的考生排序录取,第四、五、六专业志愿的考生每一档降1分与前一专业的考生排序录取。

设立专业级差的院校举例:专业级差为2分的院校的:同济大学、中国传媒大学等;专业级差为3分的院校的:吉林大学、浙江大学等。专业级差为5分的院校的:南京农业大学、南京信息工程大学、中国农业大学、湘潭大学、首都医科大学、西安电子科技大学、西华师范大学等。北京师范大学3分以内;四川大学、华北电子大学5分以内;北京第二外国语学院1-3分;南京邮电大学、南京工业大学、北京石油化工学院、东南大学1-5分以内;西北大学2-5分;西北工业大学2-8分;大连海事大学2-10分;对外经济贸易大学、北方工业大学3-5分;北京交通大学3-6分;北京大学、清华大学3-10分;西南交通大学5-10分;华东师范大学2-1-1-0;北京林业大学、北京邮电大学、中南财经政法大学为3-1-1-1;湖北大学、复旦大学、哈尔滨理工大学等为3-2-1-0;厦门大学为5-2-2-1;上海交通大学为5-4-3-0;西南大学10-5-5-0;安徽大学为15-5-5-5;武汉大学、北京中医药大学、上海财经大学为3-3-1-1-1;



用户择校因素选择

毕业省份: 湖南

毕业年份: 2012

文理科: 理科

分数: 575

城市大小	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
离家远近	<input type="radio"/> 很关心	<input type="radio"/> 一般关心	<input checked="" type="radio"/> 不关心
学校属性(是否211,985)	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
综合排名	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
师资力量, 教学排名	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
学科建设, 科研排名	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
就业前景	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
学生评价	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
生活环境	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
学习氛围	<input checked="" type="radio"/> 很关心	<input type="radio"/> 一般关心	<input type="radio"/> 不关心
分数增长速率	<input type="radio"/> 很关心	<input type="radio"/> 一般关心	<input checked="" type="radio"/> 不关心
保研比例	<input type="radio"/> 很关心	<input type="radio"/> 一般关心	<input checked="" type="radio"/> 不关心

获取推荐高校

最佳院校推荐列表

排名	学校	匹配度评分	录取平均分	志愿类型	学校优势
No.1	华东师范大学	85.93	586	冲刺型	985 211, 学校位于一线发达城市上海, 生活环境良好, 就业前景良好, 综合排名第28
No.2	四川大学	85.67	586	冲刺型	985 211, 学校位于较发达城市成都, 生活环境极好, 就业前景良好, 综合排名第14
No.3	中国农业大学	85.06	576	稳定型	985 211, 学校位于一线发达城市北京, 学习氛围极好, 生活环境极好, 就业前景良好, 综合排名第32
No.4	吉林大学	84.36	580	稳定型	985 211, 学校位于较发达城市长春, 学习氛围极好, 生活环境良好, 就业前景良好, 综合排名第11
No.5	西北工业大学	79.6	585	冲刺型	985 211, 学校位于较发达城市西安, 学习氛围极好, 生活环境良好, 就业前景良好, 综合排名第35
No.6	上海大学	78.94	577	稳定型	211, 学校位于一线发达城市上海, 生活环境良好, 就业前景良好, 综合排名第75
No.7	郑州大学	77.09	565	保底型	211, 学校位于较发达城市郑州, 生活环境良好, 就业前景良好, 综合排名第62
No.8	东北大学	76.08	572	稳定型	985 211, 学校位于较发达城市沈阳, 学习氛围极好, 生活环境良好, 就业前景良好, 综合排名第26

最佳志愿填报方案

志愿序号	学校	匹配度评分	录取平均分	志愿类型	学校优势
第1志愿	吉林大学	84.36	580	大冲刺	985 211, 学校位于较发达城市长春, 学习氛围极好, 生活环境良好, 就业前景良好, 综合排名第11
第2志愿	中国农业大学	85.06	576	小冲刺	985 211, 学校位于一线发达城市北京, 学习氛围极好, 生活环境极好, 就业前景良好, 综合排名第32
第3志愿	东北大学	76.08	572	稳妥1	985 211, 学校位于较发达城市沈阳, 学习氛围极好, 生活环境良好, 就业前景良好, 综合排名第26
第4志愿	东华大学	72.97	569	稳妥2	211, 学校位于一线发达城市上海, 生活环境极好, 就业前景良好, 综合排名第79
第5志愿	郑州大学	77.09	565	保底	211, 学校位于较发达城市郑州, 生活环境良好, 就业前景良好, 综合排名第62

收藏志愿方案

结果

(1) 近3年以内的录取分数信息很全，作为信息查询已经有了一定的规模，统一收录了各种排名和分段表信息——**可用性与价值性**

(2) 志愿推荐功能的合理性。

1. 在选择影响因素的时候，不同的选择会有不同的结果，所以基于用户不同的择校观念，同一分数会有不同的填报方案——**个性化差异！**

2. 用了近几年的20组数据进行了测试，发现系统推荐的志愿方案有极高的录取率，在第三，四志愿录取的概率达95%——**正确性！**

结论

(1) **有意义**：本系统建立在大数据上，这些数据真实，有价值，且能挖掘出更多的有用信息。

(2) **创新点**：个性化志愿推荐，具有一定的实用性，除分数外重要的决策因素，可以作为志愿填报的参考，有重要引导作用。

(3) **不足之处**：

1. 需更多更全的数据（分段表，录取人数）
2. 分数线预测，大小年
3. 专业模块(介绍，志愿推荐具体到专业)
4. 志愿信息交流等等

(4) **未来**：希望日后能获取更多数据将系统做全，并发挥它真正的实用性，能帮助广大考生进行志愿填报。

致谢

Thank you