



MyGIS 1.0 系统 设计与开发报告

GIS007 项目组

项目经理：尹伟

项目成员：吴梓杭、唐琳杰

指导老师：左泽均

2019-1-8

目 录

1	引言.....	2
1.1	项目背景.....	2
1.2	课程设计目的.....	2
1.3	课程设计任务.....	2
1.4	参考资料.....	3
1.5	定义、缩写词.....	3
2	总体设计.....	3
2.1	系统设计的原则.....	3
2.2	设计中应用的关键技术.....	3
2.3	总体结构.....	错误！未定义书签。
3	空间数据建库.....	4
4	系统功能设计.....	5
5	系统界面设计.....	6
6	模块设计与功能实现.....	7
6.1	模块汇总表.....	7
6.2	模块关系图.....	8
7	项目进程及人员分工安排.....	10
7.1	项目整体进程规划.....	10
7.2	人员分工安排.....	10
7.2.1	王晶——统筹项目进展，协调项目组成员任务安排。.....	10
7.2.2	刘德华——设计数据库，建立数据类，对数据库的数据进行管理。.....	10
7.2.3	张学友——设计与 GIS 点相关的各项操作。.....	11
7.2.4	郭富城——设计与 GIS 线相关的各项功能。.....	错误！未定义书签。
7.2.5	黎明——设计与 GIS 区相关的各项功能。.....	错误！未定义书签。
7.2.6	张艺谋——设计功能集模块的各函数及相关接口。.....	错误！未定义书签。
7.2.7	周润发——会议记录，项目相关文档的编写整理工作，项目组周报。..	错误！未定义书签。
8	实习成果展示.....	11
9	开发工具与系统运行环境.....	12
10	总结和体会.....	13

1 引言

1.1 项目背景

本系统是在地理信息系统工程课程中进行的一次综合实战演练。地理信息系统工程主要是研究：GIS 分析、设计与项目管理的相关内容。在老师的指导下，我们 3 人组成一个项目开发团队，取名为 07 GIS 项目组，选出项目经理尹伟，按照 GIS 项目开发的模式流程实施开发过程。经项目组反复讨论研究，决定开发一个武汉市出租车 GPS 轨迹处理与可视化的系统，系统取名为 07 GIS 1.0。系统主要实现对武汉市 GPS 的数据进行了可视化和数据分析处理，实现了 GPS 轨迹数据和道路网的地图匹配算法和分上下班高峰期、平峰期、深夜三个时段标注热点区域和热点路段，并从中选出了热点路段的十字路口，画出了该路口的 24 小时交通流量图和 24 小时行车速度折线图。

1.2 课程设计目的

本次地理信息系统课程设计是在完成《地理信息系统原理》、《GIS 软件工程》等课程后开展的实践性课程设计。

本次课程设计主要完成地图矢量化（空间数据库设计与建库）、系统功能设计、系统界面设计和部分功能的实现；通过这次课程设计也进一步熟悉了基于 GIS 平台提供的组件进行应用系统开发的一般方法、思路 and 流程。其主要的目的概括为：

- 1、使学生对所学 GIS 理论知识进行一次综合性复习，熟练掌握 GIS 功能与应用开发的基本流程；
- 2、培养学生对实际问题的分析能力、迁移与开放思维的创新的能力；
- 3、培养学生独立的构思 GIS 应用的能力与开发 GIS 软件的动手能力；
- 4、培养学生撰写 GIS 软件设计课程报告与文字总结开发工作的能力。
- 5、培养学生口头汇报设计与开发工作成果的能力。

1.3 课程设计任务

任务：武汉市出租车 GPS 轨迹处理与可视化

已知武汉市 2014 年 5 月的三天的出租车轨迹数据、来自 OSM 网的武汉市道路数据，请将 GPS 轨迹数据与武汉市道路数据进行可视化，并完成以下功能：

- (1) 实现数据可视化；
- (2) 实现轨迹数据与道路网的地图匹配算法；
- (3) 分上下班高分期、平峰期、深夜三个时段标注热点区域和热点路段；
- (4) 针对某个热点路段的十字路口，从以下方面分析道路服务能力：
 - 1) 24 小时交通流量图；
 - 2) 24 小时行车速度折线图；

测试数据：wuhan_GPS3Day.zip、wuhan_china.imposm-shapefiles.rar

1.4 参考资料

数据库系统概论（第四版）王珊 萨师煊 编著

数据结构、算法与应用---C++语言描述 （美）Sartaj Sahni 著 汪诗林 孙晓东 等译

计算机图形学（第三版）， Donald Hearn,M.Pauline Baker 著， 北京：电子工业出版社，2004

计算机图形学实习指导（第二版）， 郭际元、黄晓萍、曾文、龚君芳， 中国地质大学（武汉）信息工程学院，2005

GIS 分析、设计与项目管理 孙云峰 林琿 著

中地标准软件过程系列文档

1.5 定义、缩写词

DB: 英文全称 Database，数据库，是依照某种数据模型组织起来并存放二级存储器中的数据集合。

Flask: Flask 是一个使用 Python 编写的轻量级 Web 应用框架，能够根据需求开发出满足用户条件的 web 应用出来。

Echarts: ECharts 提供了常规的折线图、柱状图、散点图、饼图、K 线图，用于统计的盒形图，用于地理数据可视化的地图、热力图、线图，用于关系数据可视化的关系图、treemap、旭日图，多维数据可视化的平行坐标，还有用于 BI 的漏斗图，仪表盘，并且支持图与图之间的混搭。

2 总体设计

2.1 系统设计的原则

项目组尽可能开发出一个比较成型的 GIS 功能平台，应用所学技术使其优化，系统设计的首要原则为简单易用，效率尽可能优化，技术比较全面。

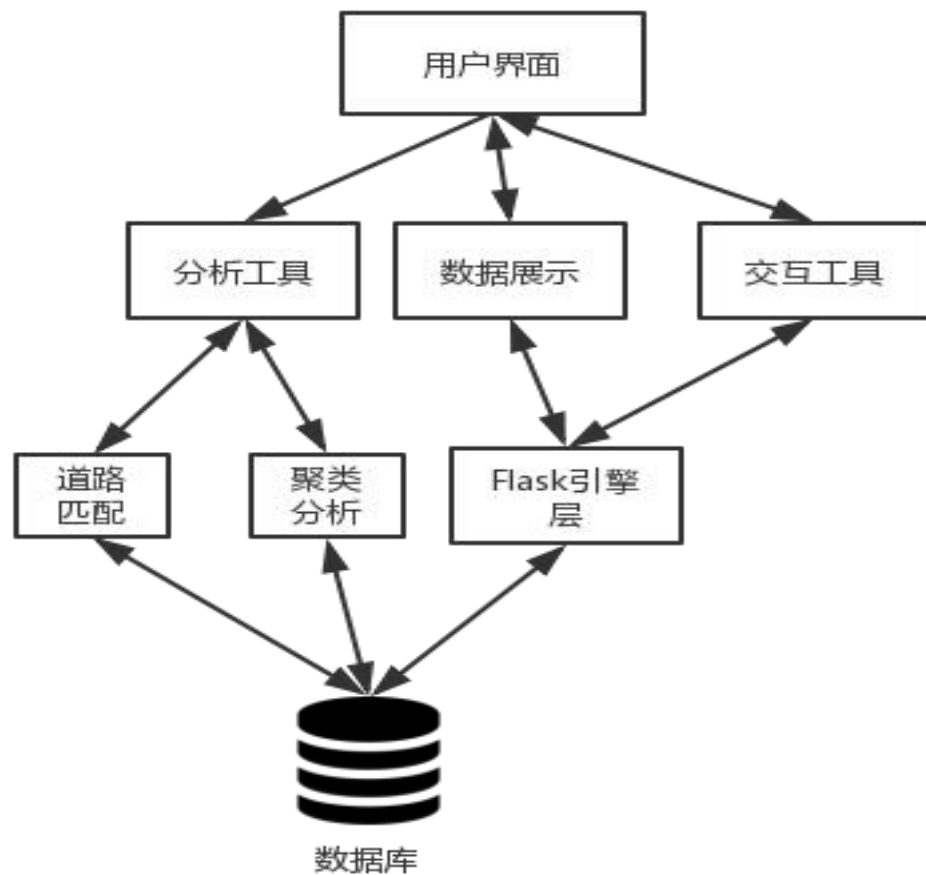
2.2 设计中应用的关键技术

由于项目组成员知识技术能力有限，缺乏项目开发经验，应用的关键技术比较少，都是课堂内外学到的知识，具体来讲主要使用了以下的关键技术。

- （1）数据处理：由于给定的数据存在数据不准的情况，所以我们需要对数据进行清洗，具体使用的工具使用 python 的 pandas 模块对数据进行清洗。
- （2）读取文件 shp 文件：这一部分要阅读官方文档中的 SAMPLE CODE
- （3）匹配武汉市道路：对每一个出租车轨迹点找到离它最近的要素，再获得要素中离它最近的 Vertex，然后用这个 Vertex 的坐标替换轨迹点的坐标。
- （4）数据聚类：使用聚类的方法找到热点区域和热点路段，然后通过百度地图的 API 可以得到该热点路段或者热点区域的经纬度。
- （5）数据库的设计：首先是经度和纬度数据类型的设计，是使用 decimal 这种类型的设计，如果使用 varchar 来设计也可，查询的时候需要用字符串匹配

- (6) 千万级别的数据库的查询和优化：由于数据库数据量上千万条，所以查询必须要优化，本次我是用的是视图和建立索引的方式来给数据库查询优化的。优化后的数据库查询由原来的 15s 左右优化到 5s 左右，缩短时间 2/3。

2.3 总体结构



3 空间数据建库

- (1) 数据库字段的设计如下

列名	数据类型	允许Null值
编号	Varchar(50)	Yes
时间	Varchar(50)	Yes
经度	Decimal(10,7)	Yes
维度	Decimal(10,7)	Yes
方向	Varchar(50)	Yes
速度	Float	No
空调	Varchar(50)	Yes

载人	Varchar(50)	Yes
----	-------------	-----

(2) 数据库的视图

由于视图是把经常需要查询的数据先查出来放到一张虚拟表中,这样就可以加快查询速率。由于第四问需要对24小时内每个小时的交通流量进行分析,所以可以把每一个小时时间段内的记录建立视图查询。其中一个例子为:

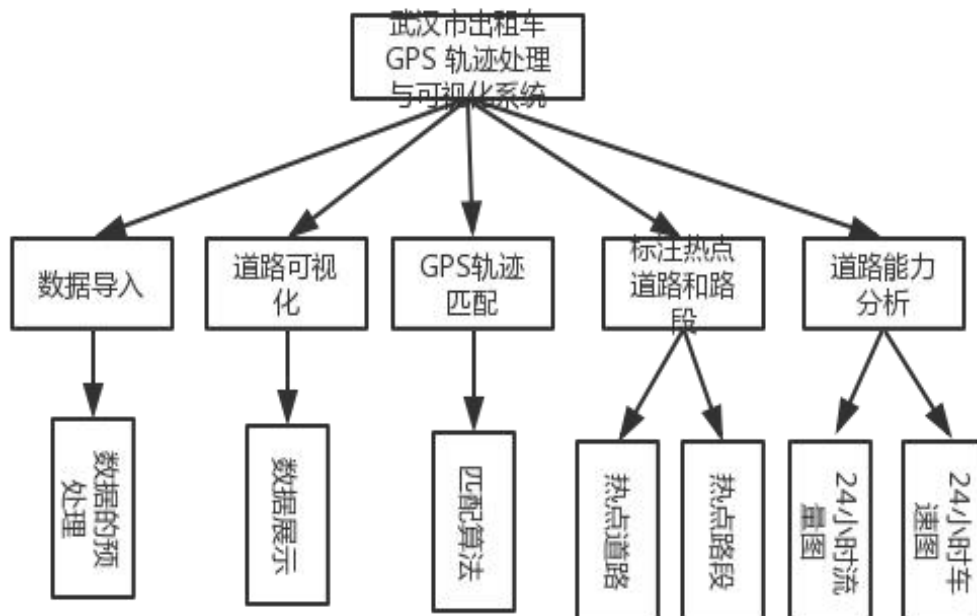
```
create view liuliangtu20140507_1
as
select * from taxi0507
where 经度>114.261 and 经度<114.289
and 纬度>30.561 and 纬度<30.589
and 时间>'2014-05-07 00:00:00' and 时间<'2014-05-07 01:00:00' and 速度>0
```

(3) 数据库的索引

为了加快查询速率,所以可以在经度纬度和时间上建立索引,提高查询速度。

4 系统功能设计

系统采用B/S和C/S混合架构的方式来构建,总体架构图如下。



首先系统总体根据题目要求分为四大部分,分别是数据可视化层、路网匹配层、热点标注层和道路能力分析层。其中数据的可视化和路网的匹配采用B/S架构,热点标注层和道路能力分析层采用的是三层B/S架构+MVC架构来完成的。下面详细介绍这些模块。

数据导入：主要是数据的处理，把数据处理成我们需要的格式，然后在导入SQL Server数据库，对数据进行操作。

道路可视化：根据数据库中的数据信息，把点打印在地图上，实现数据的可视化。展示出数据在地图上的位置。

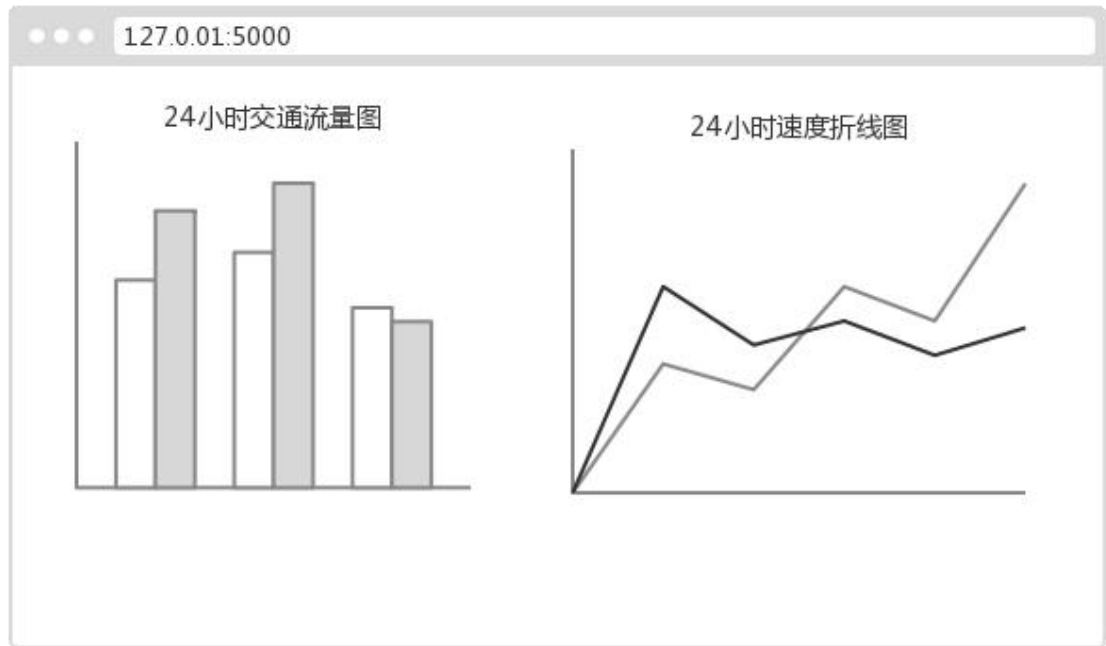
GPS轨迹匹配：由于老师给我们的数据中GPS数据不能完全匹配上路网信息，所以我们需要设计数据匹配算法，获取每一个车辆轨迹的经纬度坐标，从道路图层中获得所有线要素，对每一条线要素进行比较点到线的距离，距离点最近的线，找到其垂足即为更新的点并将数据匹配到路网上。

标注热点道路和路段：我们需要找到上下班高峰期、平峰期和深夜三个时段标注热点区域和路段，这里可以使用聚类的方法实现，然后通过调用百度地图的API可以将热点区域和热点路段找到。

道路能力分析：主要有两点指标，一是24小时交通流量图，二是24小时行车速度折线图。通过查询数据库内24小时时间段内的行车记录和车辆总数，统计得出24小时记录图。分析道路服务能力。

5 系统界面设计



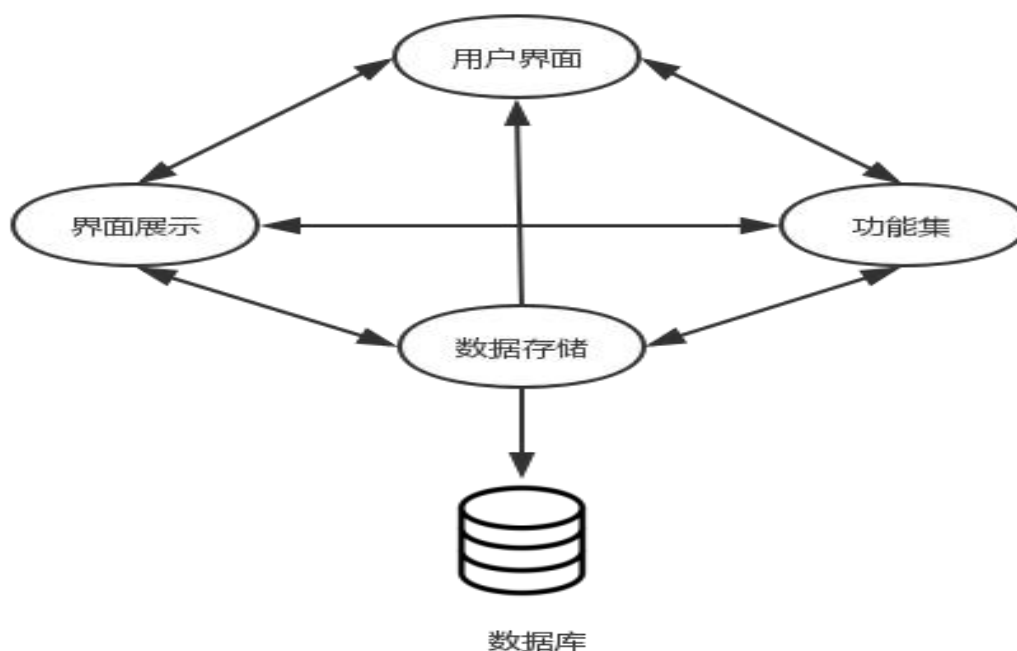


6 模块设计与功能实现

6.1 模块汇总表

模块名称	功能简述
数据处理	去除脏数据
数据展示	将数据库中的数据进行可视化，打印到地图中去
GPS 轨迹匹配	将数据库中的数据匹配到实际路网中，对于错误的数据要设计正确的匹配算法来完成。
标注热点道路	对热点道路进行标注
标注热点路段	对热点路段进行标注
24 小时流量图	统计分析某十字路口 24 小时的流量图
24 小时行车速度折线图	统计分析某十字路口 24 小时的速度折线图

6.2 模块关系图



主要功能实现的技术思路及部分源代码：

- (1) 数据的预处理，按照GB18030编码标准逐行读入原始数据，将经纬度不符合武汉市经纬度的点去除点，将错误的数据的顺序调整正确，对于每一行的字符串切割，统一重整属性的顺序和数量，写入CSV文件，将正确的文件导入数据库。

```
wuhan_cars_csv = pd.read_csv('D:/GIS/wuhan_GPS3Day/20140507.csv',
encoding = 'gb18030', delimiter="\t", header=None)[0].tolist()
wuhan_cars = pd.DataFrame()
cac = [[], [], [], [], [], [], [], []]
locate = 0
for value in wuhan_cars_csv:
    value_list = value.split(',')
    cac[0].append(value_list[0])
    cac[1].append(value_list[1])
    cac[2].append(value_list[2])
    cac[3].append(value_list[3])
    cac[4].append(value_list[4])
    cac[5].append(value_list[5])
    if 'ACC开' in value_list:
        cac[6].append(1)
    else:
```

```
        cac[6].append(0)
    if '空车' in value_list:
        cac[7].append(0)
    else:
        cac[7].append(1)
    locate += 1

    wuhan_cars.to_csv('taxi0507.csv', header=0, index=0)

(2) 通过查询数据库, 统计得出24小时内的数据
create view liuliangtu20140507_1
as
select * from taxi0507
where 经度>114.261 and 经度<114.289
and 纬度>30.561 and 纬度<30.589
and 时间>'2014-05-07 00:00:00' and 时间<'2014-05-07 01:00:00' and 速度>0

-----

create view liuliangtu20140507_2
as
select * from taxi0507
where 经度>114.261 and 经度<114.289
and 纬度>30.561 and 纬度<30.589
and 时间>'2014-05-07 01:00:00' and 时间<'2014-05-07 02:00:00' and 速度>0

select AVG(速度) from taxi0507
where 经度>114.261 and 经度<114.289
and 纬度>30.561 and 纬度<30.589
and 时间>'2014-05-07 00:00:00' and 时间<'2014-05-07 23:59:59' and 速度>0

-----

create view sudu20140507_1
as
select AVG(速度) 平均速度 from taxi0507
where 经度>114.261 and 经度<114.289
and 纬度>30.561 and 纬度<30.589
and 时间>'2014-05-07 00:00:00' and 时间<'2014-05-07 01:00:00' and 速度>0

for i in range(24):
    sql = liuliang[i]
    print('-----')
    print(sql)
    cursor.execute(sql)
    print('-----')
    row = cursor.fetchone()
```

```
print(row[0])
result.append(row[0])
remember.append(row[0])
```

7 项目进程及人员分工安排

7.1 项目整体进程规划

第一次实习：项目组成员成立，确定选题、平台、分工、编程语言、工具

第二次实习：搭建开发环境，清洗数据，按照分工进行开发

第三次实习：系统开发及测试

第四次实习：系统开发及测试

第五次实习：整理代码、文档、工程项目提交解题报告

7.2 人员分工安排

7.2.1 尹伟——协调项目组成员任务安排，设计数据库，建立数据类，对数据库的数据进行管理

第一次实习：确定开发语言和整个项目的开发思路，对整个项目有一个基本的了解，对项目人员进行分工，确定基本的项目开发进度。

第二次实习：设计数据库，建立数据库类，并将数据导入数据库，检验前两题的结论是否正确

第三次实习：设计优化查询数据库的方法，完善数据库的设计

第四次实习：完成数据可视化的所有功能

第五次实习：对所有代码、文档、工程项目进行整理，验收结题。

7.2.2 唐琳杰——完成一二两题。

第一次实习：ArcGIS for java 开发环境搭建；查看 ArcGIS for java 开发文档，对开发环境有一个基本认识。

第二次实习：查看官方文档中的相应功能的接口，可视化武汉市道路数据和出租车轨迹数据；完成对轨迹数据和道路网匹配算法的设计，绘制 UML 图，编码实现匹配算法并可视化匹配结果；组织对算法进行测试，该过程中加强与其他项目组的交流，对算法进行优化，统一规范各项资料，准备汇报。

第三次实习：对项目进行完善，准备汇报 PPT；向老师汇报这几次实习完成的情况，并展示完成的结果。

第四次实习：对匹配算法进行优化，对项目中不足的地方进行改善，最后编写实习报告。

第五次实习：验收，对所有代码、文档、工程项目进行整理，验收结题。

7.2.3 吴梓杭——设计并完成第三题。

第一次实习：开发环境的搭建，确定基本开发路线和方向。

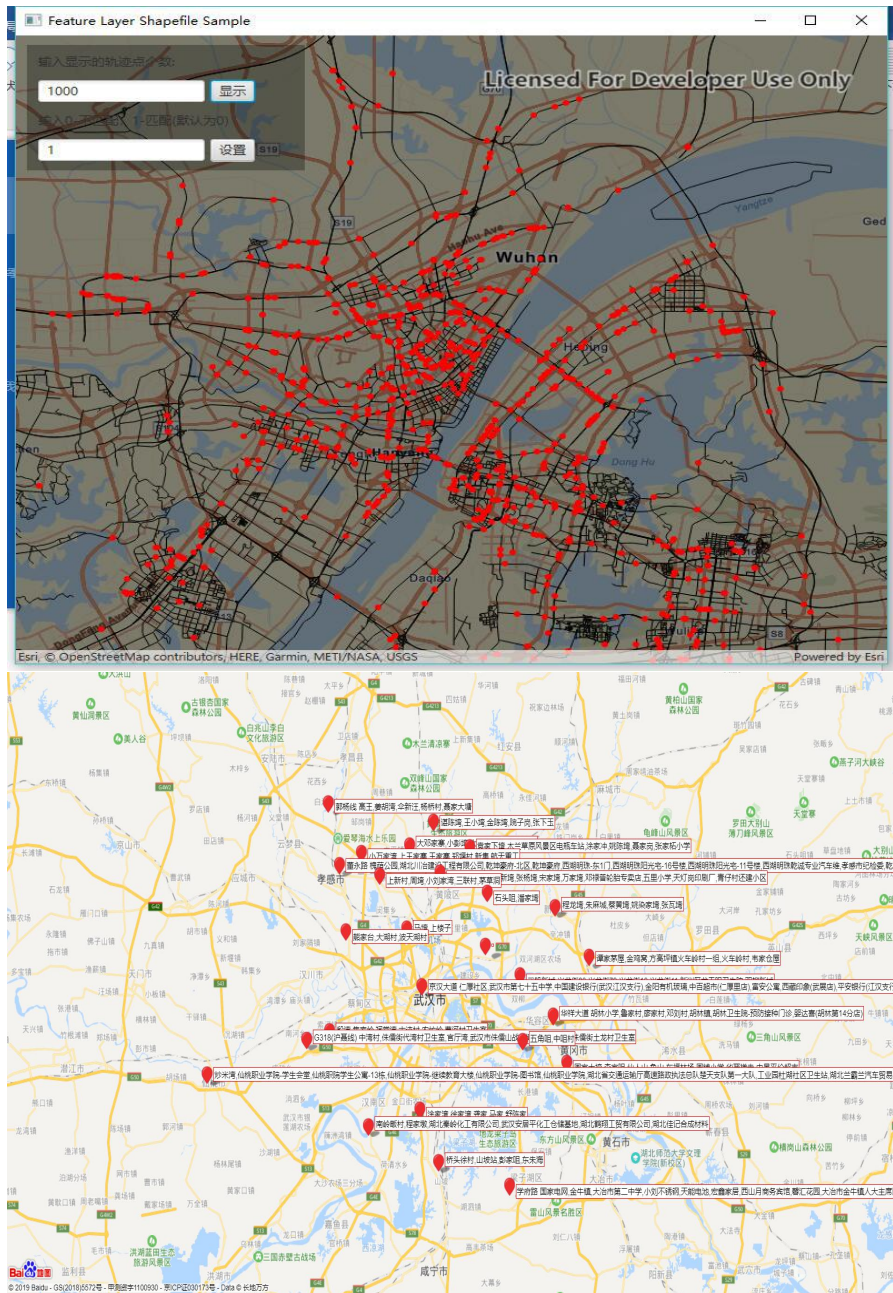
第二次实习：查看百度地图 API 并设计出初级的开发方法，并进行尝试。

第三次实习：完成整个系统第二部分的全部内容。

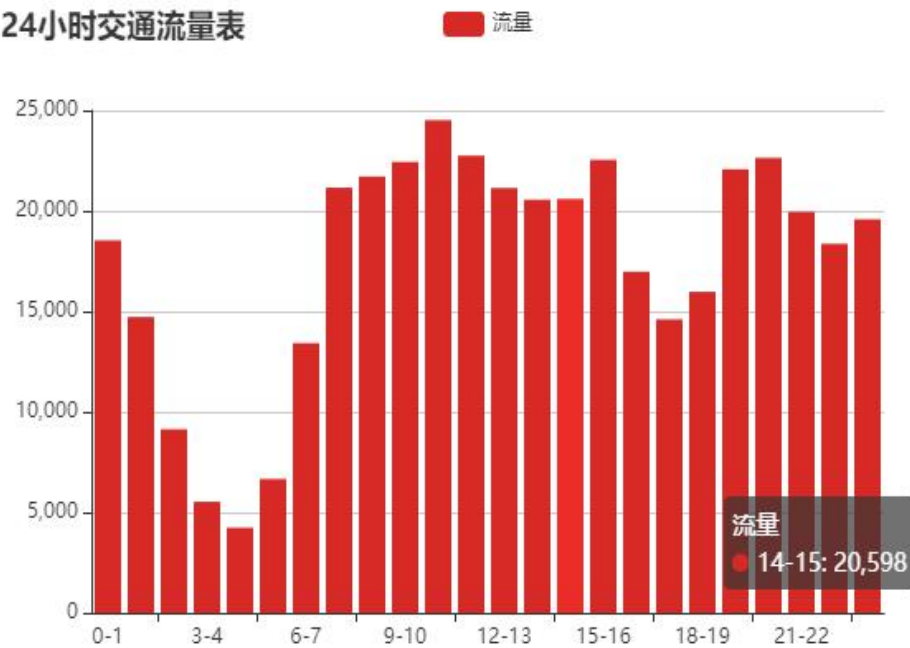
第四次实习：设计完善和优化整个系统。

第五次实习：验收，对所有代码、文档、工程项目进行整理，验收结题。

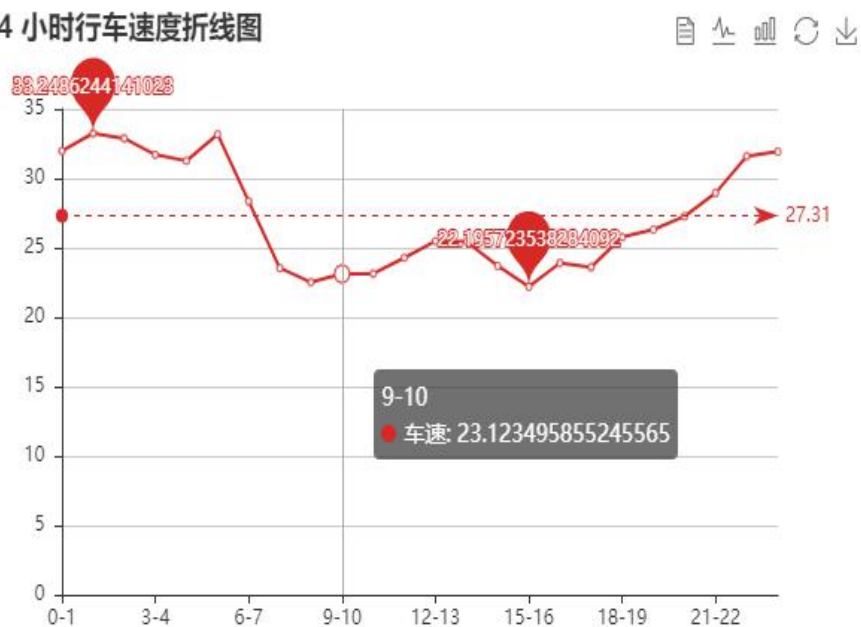
8 实习成果展示



24小时交通流量表



24 小时行车速度折线图



9 开发工具与系统运行环境

Java Idea 、pycharm 及 SQL Server 2012。

客户端操作系统：Windows 10 及其以上操作系统。

数据库服务器系统：SQL Server 2012。

10 总结和体会

(1) 通过这次实习，我深刻体会到了GIS整个系统的设计与开发的艰难，再本次开发过程中，我所做的工作主要是数据库的设计，但即便是这样看起来非常简单的工作，我任然花费了大量的时间重复设计，直到最后整个系统完成，所以通过这次实习，我深刻理解了老师上课讲的国产GIS系统的苦衷，我想这样一批默默无闻发展国产GIS系统的老师和教授们致以诚挚的敬意，我对他们所做的工作深感敬佩，我觉得在信工学院，我最大的收获不仅仅是知识的增长，更是我看到了一群为祖国地理信息产业默默奉献的一群人，他们也许不被当下主流媒体所报道，但是他们所做的工作绝对是富有意义的，通过这次实习，我个人的出路也许多了一条，我也可以去开发GIS相关的系统，而不是在软件工程一棵树上吊死。

(2) 通过这次实习加深了我对课堂知识的了解，充分地使用到了以前学到过的知识，我觉得这次实习是一次非常好的实习，锻炼了同学们的能力，让同学们和我体会到了GIS系统的开发过程中。

(3) 在完成第四问的过程中，一开始查询速度很慢，这时我想到数据库中可以进行查询优化的方法，于是便通过建立视图、索引等工具实现了加速查询。

(4) 但是我也觉得这堂课还是需要完善的地方，一是课程实习的时间安排上不够合理，一个星期完成本次实习时间上稍显紧张，如果真的只有这么短的时间安排实习，可以提前发题。二是课堂内容还可以更加贴近专业化，软工的同学不是不愿意学GIS，老师可以通过考试等方式来提高同学们的参与率，在课堂上讲述专业知识也是非常必要的。另外我的建议就是上课的时候可以通过及时出练习题给同学们做，加深同学们对知识的理解，在一个，就是可以设立学生讲堂，因为老师和学生看问题的方法往往不一样，而学生讲过的东西学生容易理解，没有代沟。