



# 数据库专题训练

数据库新型检索技术

大作业 互联网出行

助教 李浩达 [lihaoda9@163.com](mailto:lihaoda9@163.com)



# 大作业说明



- 互联网拼车情景设定：
  - 给定10万出租车
    - 当时经纬度
    - 乘客上限 – 4个
    - 车上乘客终点
  - 给定1个查询——乘客当前位置
    - 找到最合适的5个车（不超过上限）
    - 出租车绕路不超过10km
    - 出租车离乘客不超过10km
  - 请自行选择编程语言、框架，实现程序完成上述情景下的查询



# 大作业说明



- 为了简化问题，我们假设：

- 乘客：

- 乘客均独自一人，仅占用车上一个位置，有一个终点
- 不考虑乘客下车后空出来的位置

- 出租车：

- 所有出租车上均有4个位置
- 出租车接单后将立刻去接乘客，而不会先运送车上乘客
- 出租车将沿最短路径前进，不考虑拥堵、掉头等情况

- 路网：

- 出租车和待接乘客都位于路网节点上，不会处于边（道路）中间



# 大作业说明



- 绕路定义:

- 假定当前车上有 $k$ 个乘客，依次送达目的地所需路程为 $D1$ ，出租车接乘客上车所需路程为 $D2$ ，接到乘客后将所有 $k+1$ 个乘客送达目的地所需路程为 $D3$ ；待接乘客位置到自己目的地的最短路距离为 $D4$ 。
  - 车上乘客绕路:  $D2+D3-D1$
  - 待接乘客绕路:  $D3-D4$
  - 要求上述两个绕路距离均不能超过10km





# 大作业说明



- 数据说明

- 路网数据

- road.cnode
    - 每行代表一个节点的位置，包含3个数字，分别为节点编号、经度、纬度，以空格分隔
  - road.nedge
    - 第一行2个数字，分别表示节点数、边数。
    - 接下来每行代表一条边，包含3个数组，前两个为节点编号，最后一个为道路长度，以空格分隔。
    - 所有边均为无向边，可双向通行。



# 大作业说明



- 数据说明

- 出租车数据

- car.txt

- 每行代表一辆出租车，以空格分隔包含3+k条数据，分别为出租车编号、车上乘客数目、出租车位置、乘客目的地位置
      - 出租车位置和乘客目的地位置均以逗号分隔，含有3个数字，分别表示经度、纬度、路网上的节点编号



# 问题简化



- 出于简化问题考虑，各位同学可以不使用路网，仅使用经纬度坐标，依据欧式距离来完成大作业
  - 若不使用路网，则绕路距离、出租车与乘客距离均以欧式距离来计算
  - 若使用路网，则绕路距离、出租车与乘客距离均以路网上的最短路径距离计算
    - 可以使用第三方寻路算法库，比如  
<https://github.com/TsinghuaDatabaseGroup/GTree>



# 评分标准



- 输入一个乘客位置，在可以忍受的时间内(5-10秒)，返回不超过5个有空位的出租车。所有返回的出租车与乘客距离不超过10km，若没有合适的出租车，则返回空列表【60%】
- 返回的所有出租车的绕路距离不超过10km【20%】
- 提供一个UI界面方便输入输出的交互，包括设定、显示乘客位置，显示返回的出租车位置、车上乘客目的地等。建议借助一些第三方资源完成。（比如百度地图什么的）【10%】
- 使用路网数据完成大作业。【10%】
- 加分项：返回结果中包含建议的车辆行进路线、乘客送达顺序，结果合理并显示在UI中。【10%】
- 加分项：自行实现路网寻路算法。【10%】





# 提交方式



- 在网络学堂提交书面材料，包括：
  - 源代码
  - 程序编译/运行说明
    - 说明程序如何编译、如何运行、数据文件如何存放等等。
  - 大作业设计说明/报告
    - 说明大作业的程序设计框架、算法、使用过的第三方库等。
- 现场演示：
  - 为助教现场展示大作业中所实现的各个功能



# 截止日期



- 书面材料
  - 第十五周周末，6月7日
  - 以网络学堂为准
- 现场演示
  - 时间(暂定)
    - 第十四周 周四、周五（5月30日、5月31日）
    - 下午2点-5点
  - 地点
    - 东主楼 10区204
- 如有变动，将及时在网络学堂发布通知
- 如有同学无法在这段时间来检查，可以给助教发邮件预约时间



# Thanks, Questions?