



## 城市导航迷路场景关联知识图谱与计算系统

项目负责人： 李安南

项目组成员及联系方式：

李安南（17814562133） 李思琪（18788838860）

方天桐（18071100505） 李梦萍（15070473397）

指导老师：杨林 （联系方式 15927322600）

中国地质大学（武汉）

2023 年 6 月

# 目 录

一、前言 .....	2
二、项目概述 .....	2
1. 项目背景及意义 .....	2
2. 项目目标及内容 .....	2
3. 项目研究基础 .....	3
4. 项目应用前景 .....	3
三、产品介绍 .....	4
1. 产品概述 .....	4
2. 产品特色 .....	4
3. 产品功能 .....	4
四、组织设计.....	11
1. 团队分工.....	11
2. 对知识和工具的掌握.....	12
五、项目实施.....	13
1. 具体实施内容.....	13
2. 模型与系统设计.....	14
3. 方案可行性分析.....	15
4. 环境分析.....	15
六、项目展望 .....	16
1. 国内外研究现状 .....	16
2. 项目特色及创新点 .....	17

## 一、前言

本项目旨在探索和分析用户在使用导航软件时遇到的问题，以及为导航软件提供改进方案，从而提升用户的体验。

导航决策是人们日常出行的一项重要认知任务，导航经验知识的获取对于人类导航行为及导航能力至关重要。然而现有大数据环境下日益丰富的导航经验知识并不能由现有导航地图覆盖或表征，远未形成激发导航者认知潜能提升的有效学习通道。因此，本项目针对导航决策中容易出现的迷路场景，以群集导航经验文本为数据集，基于深度学习方法构建面向导航迷路场景关联知识图谱与导航迷路场景的查询计算方法，以揭示导航共性经验规律，服务于大众的导航空间知识获取与自主导航决策行为的效能提升。

## 二、项目概述

### 1. 项目背景及意义

导航决策与人们的日常出行息息相关，是人们日常活动中的一项重要的认知任务。导航经验知识的获取对于人类导航行为及导航能力至关重要。随着城市化进程的不断加速，城市道路和城市环境的复杂性日益加剧，迷路成为了人们在陌生环境下经常会遇到的问题。因此，本项目着重针对导航决策中的迷路场景进行特定的经验知识图谱与计算研究，并提供智能搜索和迷路经验自然语言智能问答。使得导航者在面临实际复杂交通环境及迷路场景时，具备独立于导航系统辅助之外的自主导航意识，以及应对潜在变化与解决冲突的能力。

此外，现有大数据环境下产生了日益丰富的导航经验数据，这些数据在导航过程中借由用户与环境交互形成，蕴含着丰富的导航认知线索，代表了典型用户群体的共性导航经验认知。这些经验知识可以增强导航者导航决策的质量，对大众导航认知水平的提升具有宝贵的价值。然而，这些群集经验并不能由现有导航地图（由制图专家认知主导）完全覆盖或准确表征，加之目前经验建模分析的局限，缺乏将群集经验吸收进导航地图的手段，阻碍了它们作为高级认知经验或结构化经验在导航应用中进行共享和学习。

在此背景下，如何创新性地借助众源导航迷路经验数据学习，揭示共性经验规律，构建迷路场景经验知识图谱与计算的基础框架，增强导航系统对导航环境空间知识的获取及路径决策能力，是一个非常值得研究的重要基础科学问题。故，本项目拟开展“城市导航迷路场景关联知识图谱与计算研究”，有望形成基于更广泛群集经验认知的经验共享与学习通道，快速提升大众导航认知的层次和水平，具有重要理论意义与实用价值。

### 2. 项目目标及内容

本课题以提升导航主体认知学习能力为目标，创新性地借助众源导航迷路经验数据学习，构建导航迷路经验的关联知识图谱与查询计算方法，通过导航决策驱动语义视角下的导航迷路经验知识的挖掘与可视化呈现，打通导航者通过导航地图学习导航经验知识的学习通道，为用户提供一种新颖的探索环境知识交互学习的导航服务形式。

具体目标包括：

- 1) 导航迷路场景的知识图建模，将迷路场景的时间、空间、天气、地形、道路、建筑等环境特征等概念进行抽象与关系建模。
- 2) 导航迷路经验特征词集抽取，实现对与导航决策相关的基础导航要素与本体经验语义集的提取；
- 3) 导航迷路场景的查询计算，针对迷路场景的特定查询构建图存储方案以及查询优化策略；

4) 导航迷路场景的可视化系统构建。建立融合导航迷路经验语义的新型导航地图表征, 促进导航用户加速对复杂迷路环境的学习与理解。

主要内容:

1) 原始迷路导航经验信息管理: 收集、整合、存储、管理用户在出行过程中的导航经验信息。

2) 对导航迷路经验信息进行时空要素抽取并展示: 从导航经验信息中提取时间、天气、地点、环境因素等信息, 并将其在地图上展示。

3) 对导航迷路的场景进行分类和统计展示: 将导航经验信息中用户迷路的场景进行分类和统计, 使用户更好地了解迷路的原因和场景。

4) 对导航迷路场景的知识图进行计算、展示与查询: 将导航经验信息中用户迷路的场景构建成知识图谱, 并通过计算、展示和查询等方式, 让用户更好地了解不同场景的迷路情况。

5) 导航迷路情感分析: 对导航经验信息中用户迷路的场景进行情感分析, 以使用户更好地了解用户在迷路场景中的情感体验, 为导航软件提供改进方案。

6) 导航迷路因果关系分析: 对导航经验信息中用户迷路的场景进行因果关系分析, 以使用户更好地了解导致迷路的原因, 为导航软件提供改进方案。

### 3. 项目研究基础

本课题为国家自然科学基金项目“融合城市导航场景空间认知与群集经验学习的导航地图表征模型研究”(项目编号: No. 42071383) 中的研究内容, 项目已经具备了导航经验文本分类、实体抽取和知识图谱的相关技术基础。本项目的研究内容涉及自然语言处理、深度学习、大数据分析等多个领域, 已有相关研究成果和技术积累。近年来, 自然语言问答系统、深度学习模型在各领域得到广泛应用, 并取得了大量成果。同时, 大数据分析技术也在城市规划、交通管理等方面得到了广泛应用, 并有望在导航系统中发挥重要作用。

本项目的研究内容涉及到多个学科领域的技术, 需要具备相关领域的专业人才、硬件设备和软件工具。本项目的主要人员具备相关领域的专业知识和经验, 拥有充足的研究经费和实验设备, 能够提供足够的学术支持。项目涉及到多个学科领域的技术, 需要跨学科合作, 整合各领域的专业知识和技术。同时, 本项目涉及到大规模数据处理和分析, 需要建立完备的数据管理和分析平台。为此, 需要建立跨学科合作机制, 整合各领域的专业知识和技术, 建立完善的数据管理和分析平台, 提高数据处理和分析的效率。

### 4. 项目应用前景

1) 旅游行业: 该项目可以为游客提供更准确、更全面的导航信息, 使游客更加轻松地找到自己想去的景点和游玩路线, 提升游客的旅游体验。

2) 城市交通: 该项目可以为居民提供更准确、更实时的交通信息, 包括路况、公交地铁等信息, 让居民更加方便地出行, 缓解城市交通压力。

3) 城市规划: 该项目可以为城市规划者提供更全面的城市数据, 包括人流、交通流、商业分布等信息, 帮助规划者更好地规划城市的未来发展。

4) 商业服务: 该项目可以为商家提供更准确的商业信息, 包括人流量、消费习惯等信息, 帮助商家更好地了解消费者需求, 提升商家的服务质量和客户满意度。

5) 智慧城市: 该项目可以作为智慧城市的重要组成部分, 通过建立城市导航知识图谱, 实现城市交通、旅游、商业等多个方面的信息共享和智能化管理, 推动城市建设向智慧城市转型升级。综上所述, 该项目在旅游、交通、规划、商业和智慧城市等多个领域都有广泛的应用前景, 可以为人们的生活和工作带来更多的便利和效益。

## 三、产品介绍

### 1. 产品概述

城市导航迷路场景关联知识图谱与计算研究项目，旨在通过构建城市导航知识图谱，结合人工智能技术，提供更为准确、全面的城市导航服务，让用户更加方便地出行和生活。该项目产出的“城市导航迷路场景关联知识图谱与计算研究”原型系统，拥有用户友好的简便易用的交互界面和智能高效的导航服务。

### 2. 产品特色

1) 智能化：该产品利用人工智能技术，结合大数据分析和机器学习算法，实现导航信息的智能化处理和推荐，提供更为准确、全面的导航服务。

2) 个性化：该产品可以根据用户的出行需求和习惯，为用户提供个性化的导航方案和推荐，让用户更加方便地出行和生活。

3) 实时性：该产品可以实时监测城市交通情况和人流状况等信息，为用户提供最新、最准确的导航信息。

4) 全面性：该产品可以提供全面的城市导航服务，包括步行、公交、地铁自驾等多种出行方式，覆盖城市各个角落和场景，让用户更加方便地出行和生活。

5) 可视化：该产品可以通过地图等可视化方式，展示导航信息和路线规划，让用户更加清晰地了解导航情况和路线规划。

综上所述，该产品具有智能化、个性化、实时性、全面性和可视化等特色，可以为用户提供更为准确、便捷的城市导航服务。

### 3. 产品功能

#### 1) 迷路场景知识图谱：

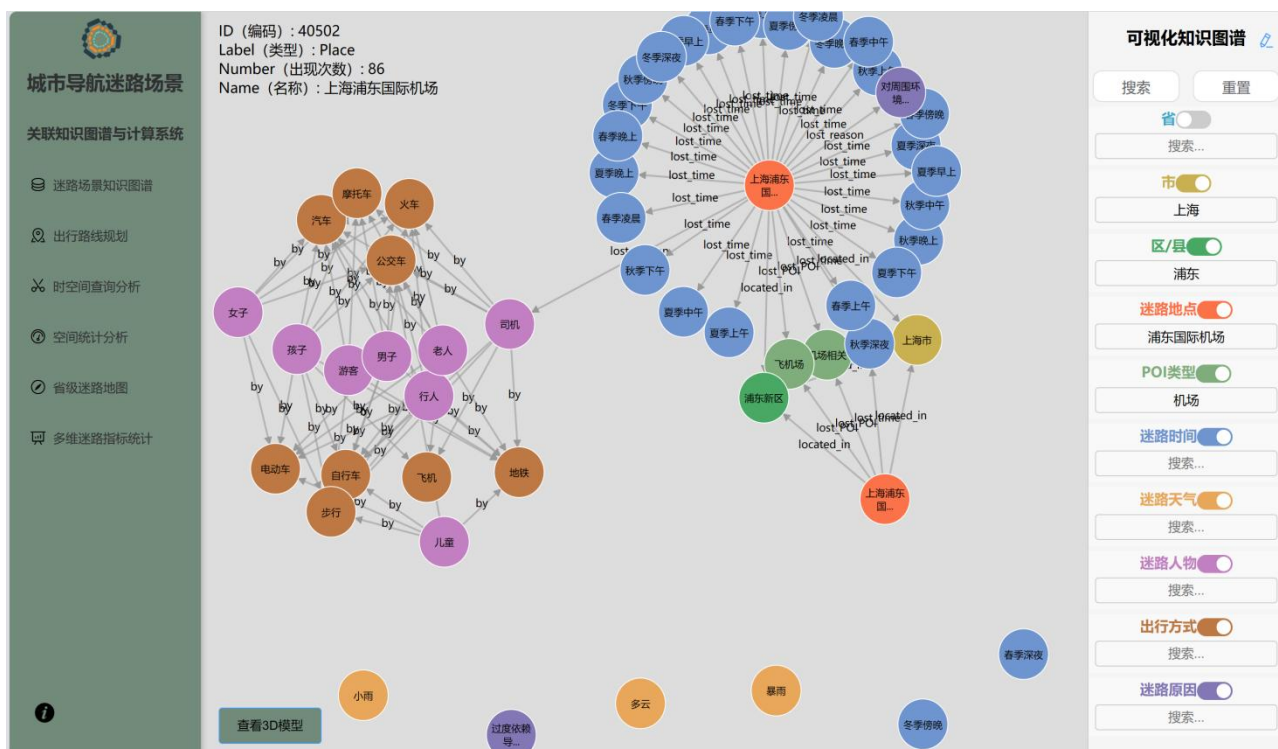
1. 支持多种图数据库中各类节点和关系数据的可视化展示以及特定的节点数据展示，用户可以通过在搜索框中输入限制词来查找特定的节点，通过打开和关闭开关来选择节点图形的可见性。

2. 重置按钮可以帮助用户清空画布和搜索框内容。

3. 为了方便用户观察，所有节点都提供有拖拽和移动功能，图形画布也可以通过拖拽移动来观察超出页面范围的溢出内容。

4. 为了使用户能观察到各个节点的详细信息，为每个所有节点添加了点击响应事件，通过点击按钮用户可以查看节点的基础信息和详细内容，包括编码、类型、出现次数和名称。

5. 查看 3D 模型按钮允许用户跳转到托管 3D 模型的第三方网站 [app.graphxr.cn](http://app.graphxr.cn)，该网站提供了多样的 3D 化图数据库展示效果和外部工具。



在该查询结果中，我们可以观察到，在冬季深夜、夏季晚上等时间节点在上海浦东国际机场容易发生迷路事件，该地点隶属于上海市浦东新区，司机群体更容易因为对周围环境不熟悉在该地点附近迷路，点击节点可以发现我们在收集到的所有数据中浦东国际机场总共出现了 86 次。此外，司机群体与飞机、地铁、火车等交通工具造成的迷路事件也有一定关联性，推测可能是在寻找火车站、地铁站的过程中容易发生迷路事件。

图 1. 查询节点和关系

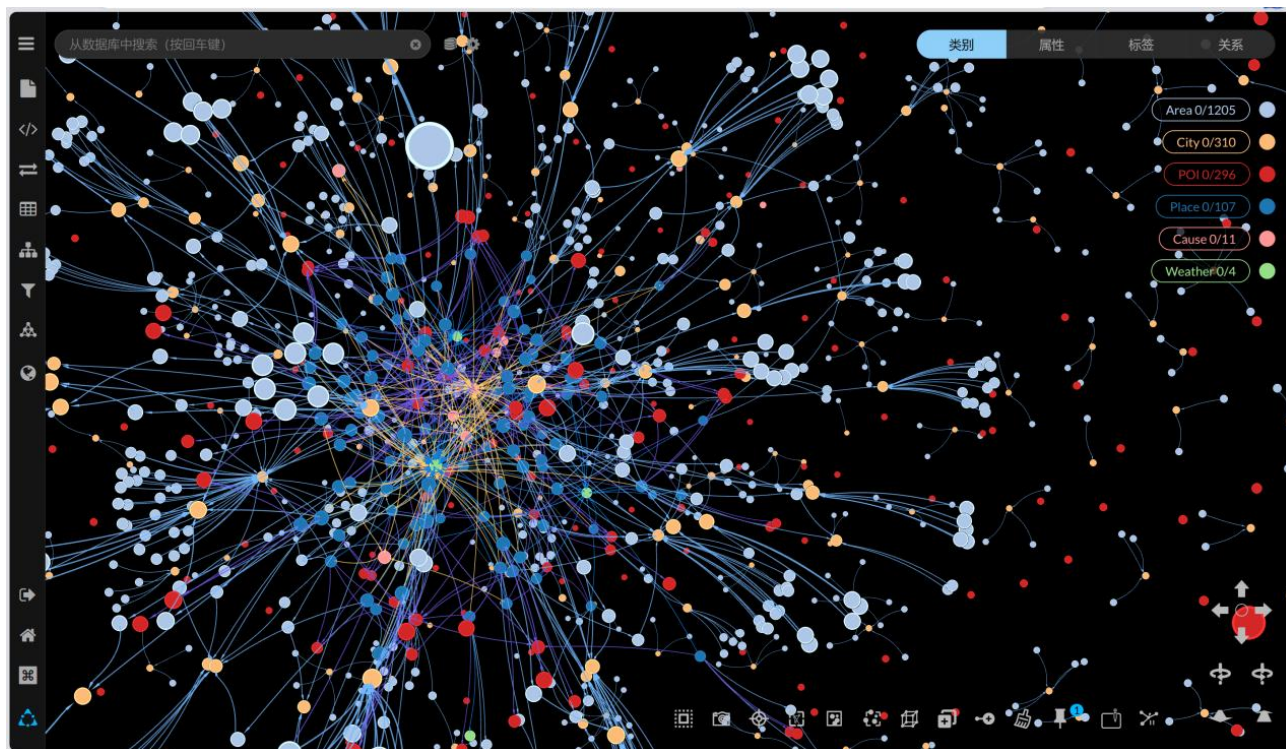
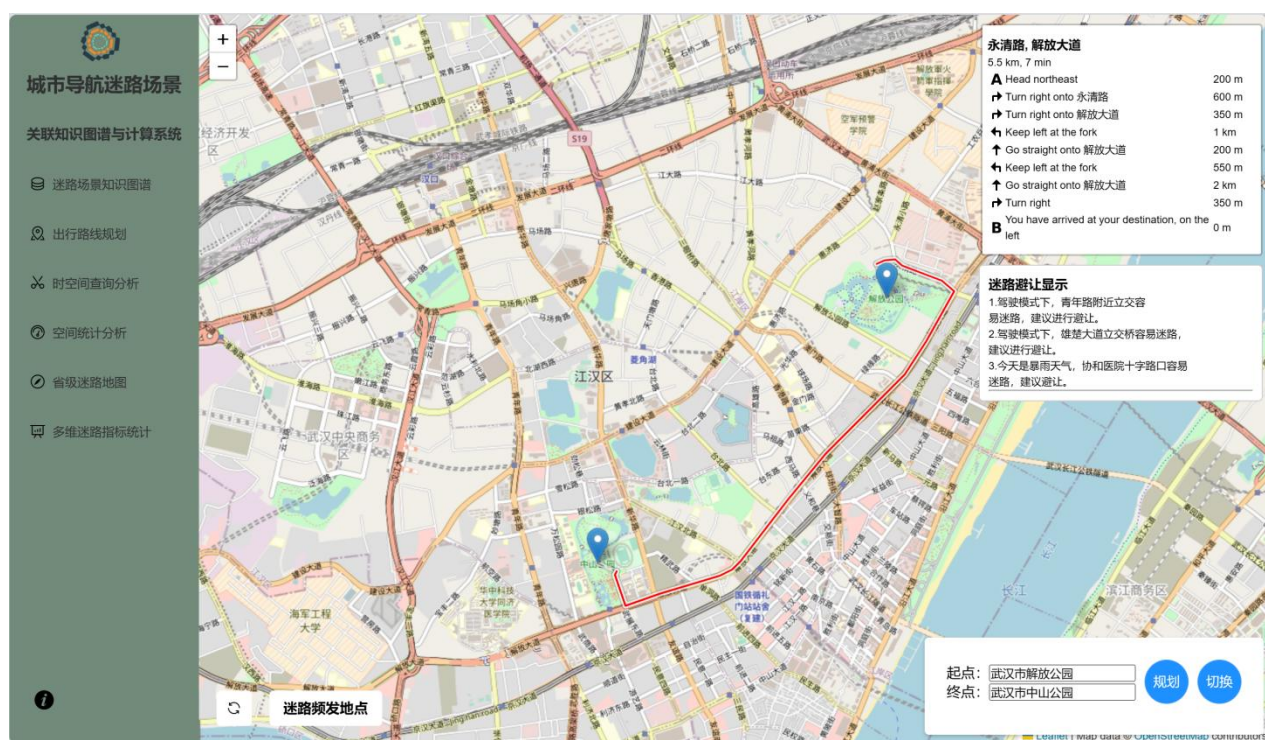


图 2. 3D 模型



## 2) 出行路线规划:

1. 用户可以通过在输入框中输入起点和终点来进行路径规划,系统地图将给出最适合的路径并在地图上绘制并在右上角描述路线的详细内容,同时提供与该路线相关的行进提示。
2. 切换按钮允许用户切换到地点标注模块,该功能通过将用户输入地名转化为经纬度并在地图上标注,并跳转到特定地点,通过右击图标可以删除标注点,左击图标可以查看标注点详细内容。
3. 迷路频发地点按钮能调用高德 API 将图数据库中的所有迷路地点名转换为经纬度,并将所有地点标注在地图上,以方便用户查看和分析。
4. 刷新按钮能帮助用户清空当前界面地图上存在的全部标注内容。



1. 系统根据用户指定的起点和终点我们规划出了预期的行进路线,该路线从武汉市解放公园起,途径永清路、解放大道到达武汉市中山公园。
2. 同时,我们根据已有数据弹出了迷路避让提示,告知用户可能潜在哪些导致迷路的因素,如驾驶模式下青年路和雄楚大道附近可能发生潜在的迷路事件、暴雨天气下协和医院十字路口可能发生潜在的迷路因素,需要避让。

图 3. 路线规划

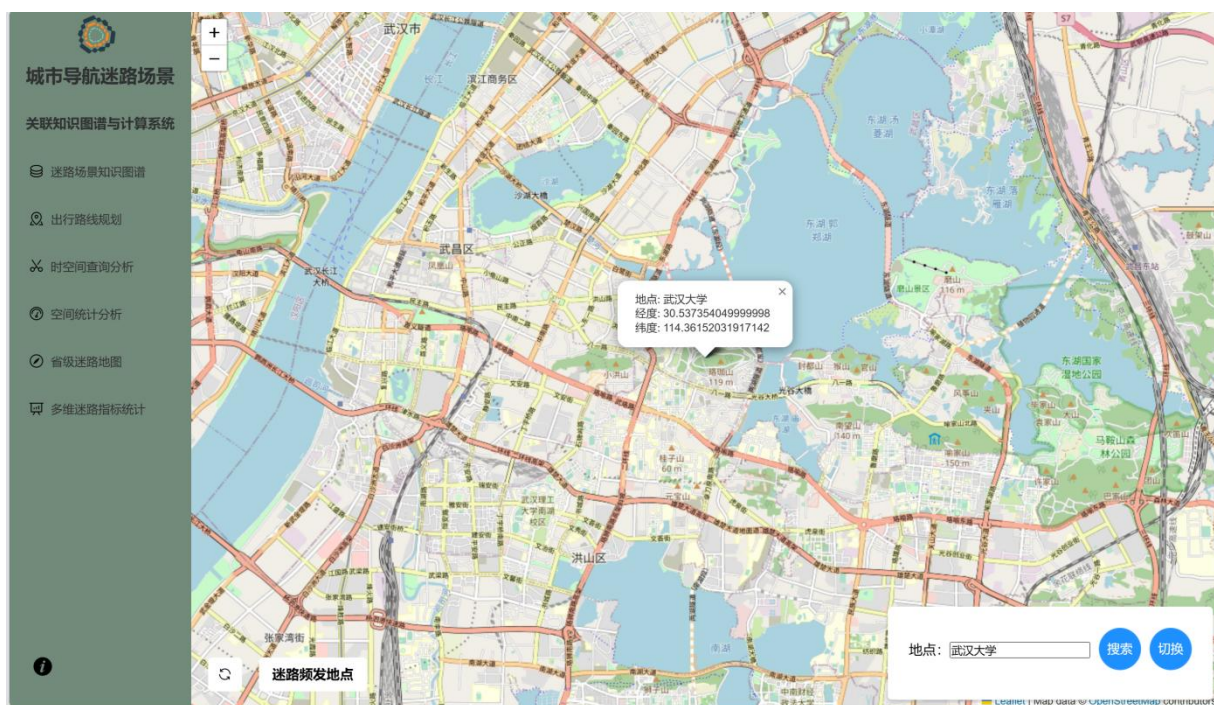


图 4. 特定地点标注

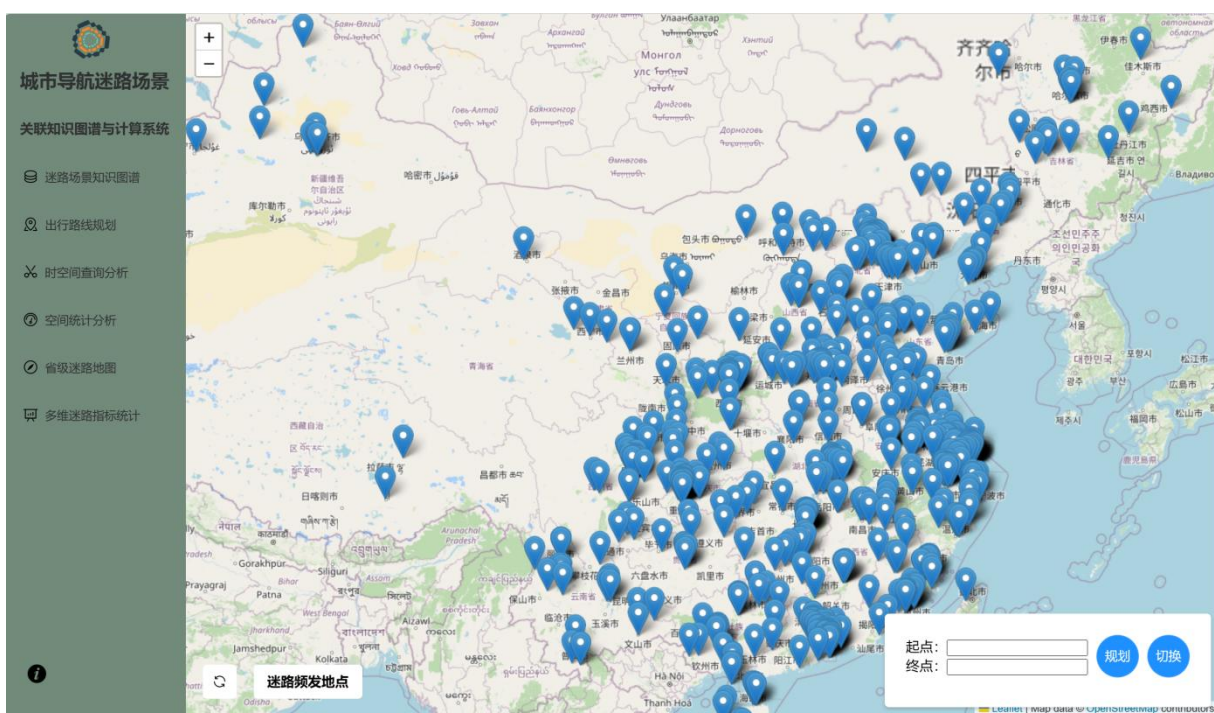


图 5. 迷路频发地点

### 3) 时空间查询分析:

1. 允许用户使用绘制按钮在地图上进行圆形范围的绘制, 能够将在用户绘制的圆形范围内的迷路地点数据标注出来, 方便用户以一种更加直观的方式来观察数据点的分布以及相对位置。
2. 右下角的查询搜索框指定城市行政区划范围内查询功能, 允许用户查询某个指定城市范围内的全部迷路地点数据。



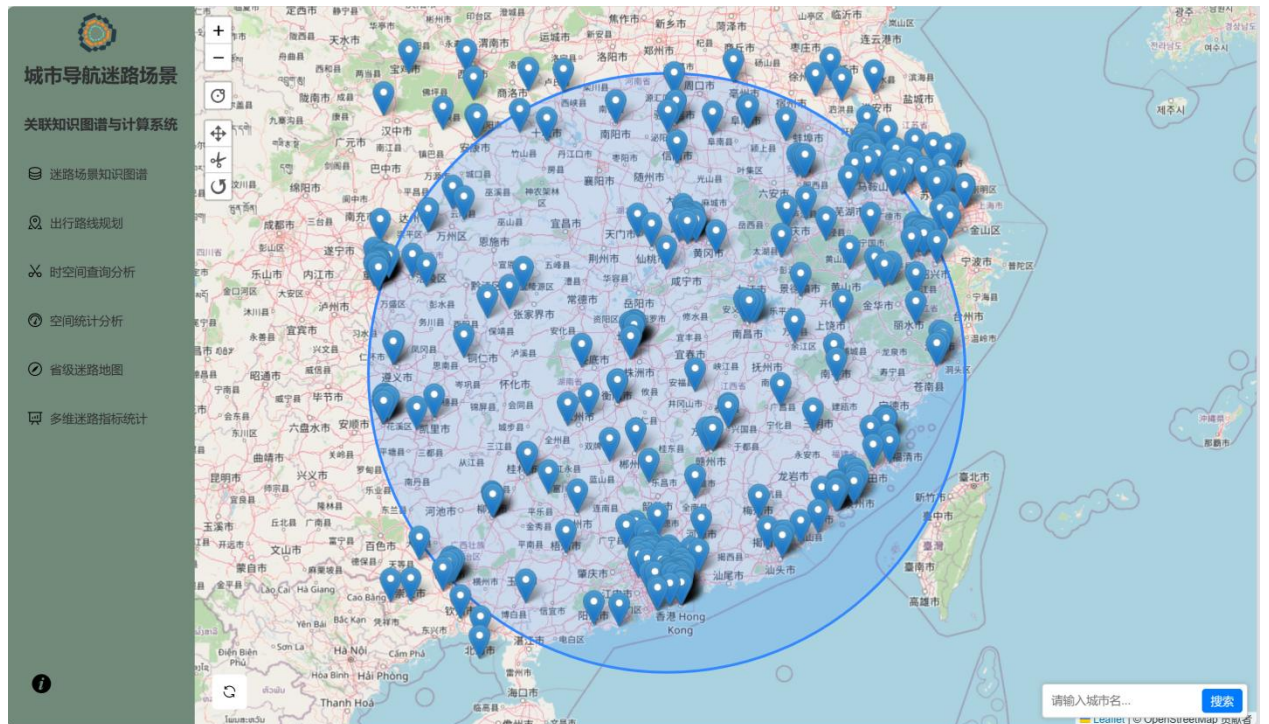


图 6. 特定范围查询

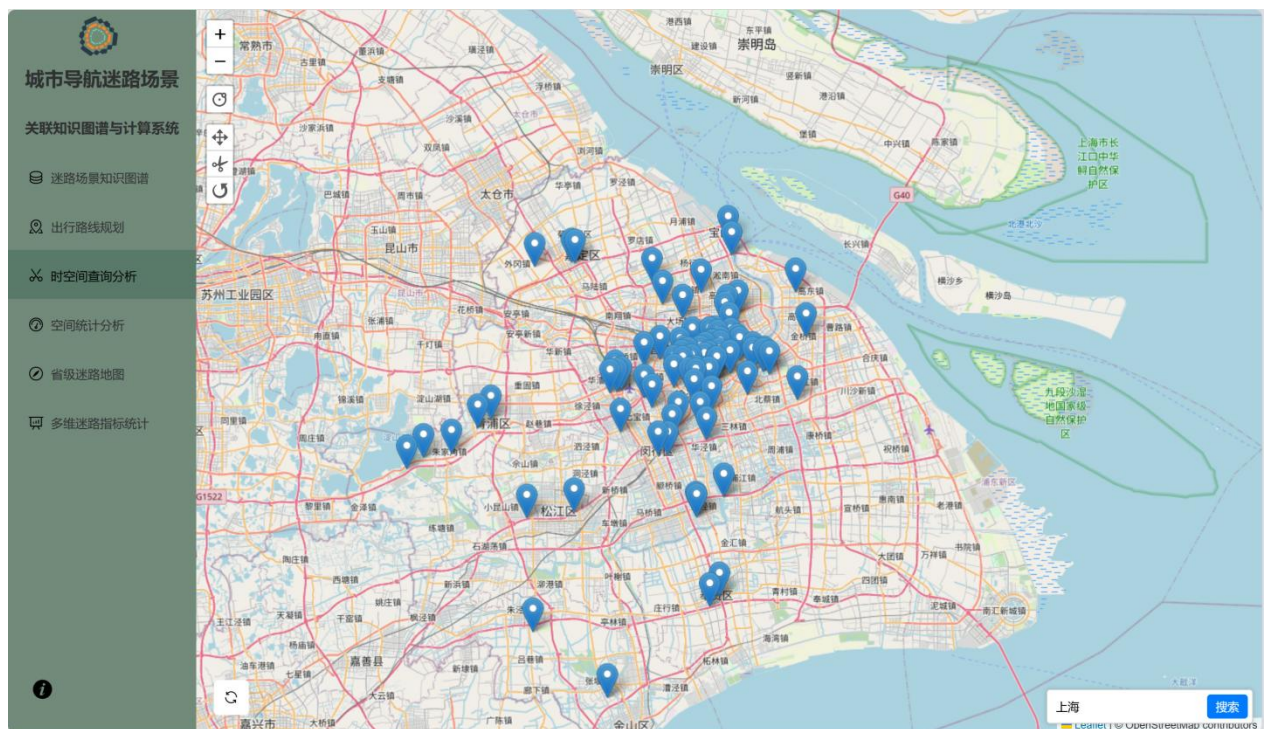
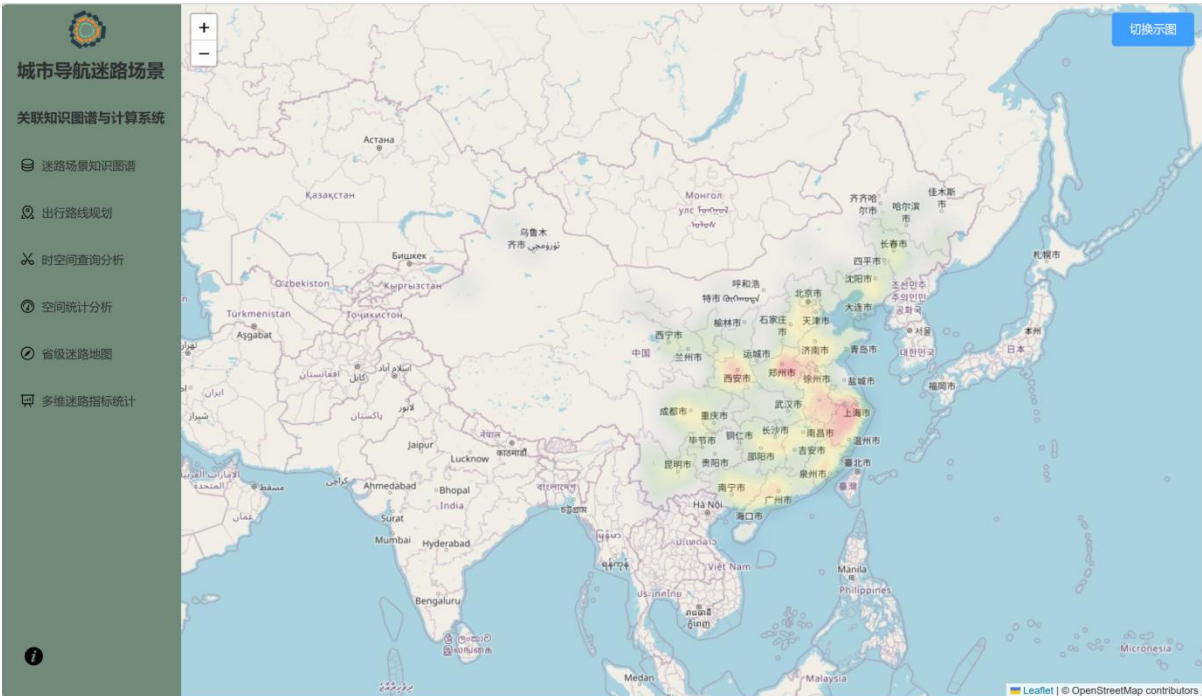


图 7. 指定城市查询

#### 4) 空间统计分析:

1. 通过调用高德地理编码将数据库中的地点转换为经纬度，使用 Leaflet API 显示 MapGis 地图在全国范围内绘制热力图，向用户展示数据的地理分布特征，直观地表现不同地区数据的密集程度和分布情况。

2. 右上角的切换视图按钮提供一个弹窗, 包含热力图所使用的迷路地点数据的柱状图和折线图, 用于辅助用户分析板块所绘制出的热力图内容。



根据图例显示, 城市级迷路数据的高频发生地带集中在上海市、郑州市、南京市、徐州市、沧州市、重庆市等城市附近, 这可能是地理位置、人口和经济因素造成的。

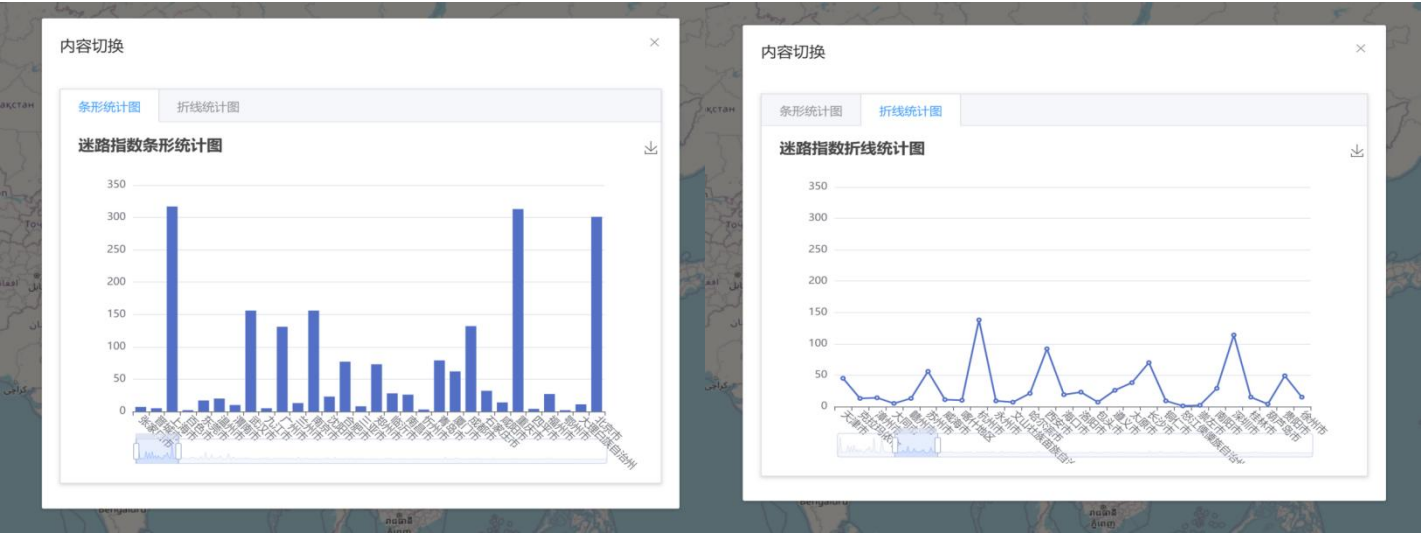


图 8. 热力图及其辅助统计图

5) 省级迷路地图:

1. 使用颜色根据数据量对各省行政区划的图形进行分级显示, 向用户提供了数据在省级行政单位尺度上的数据分布情况, 使得用户能鲜明地观察到各个省份采集到的数据多寡。
2. 通过将鼠标悬停, 用户可以观察到目标省所收集到的总数据量。





在该图中我们可以看到拥有最多迷路数据的省级单位地点是广东，共计拥有 433 条迷路相关数据。其中，江西省自身迷路数据较少周围省份迷路数据却较高，形成了“low-high”分布模式，而重庆市则属于“high-low”分布模式。

图 9. 省级迷路地图

#### 6) 多维迷路指标统计:

1. 向用户提供扇形统计图、环形统计图、条形统计图、折线统计图，允许用户得以观察各类数据在数据库中的占比和数量，进行数据统计和分析。
2. 右侧侧边栏提供了数据项的选择功能，用户可以根据需要选择特定数据项进行查看和分析，观察该类型数据在数据库中的组成和占比。

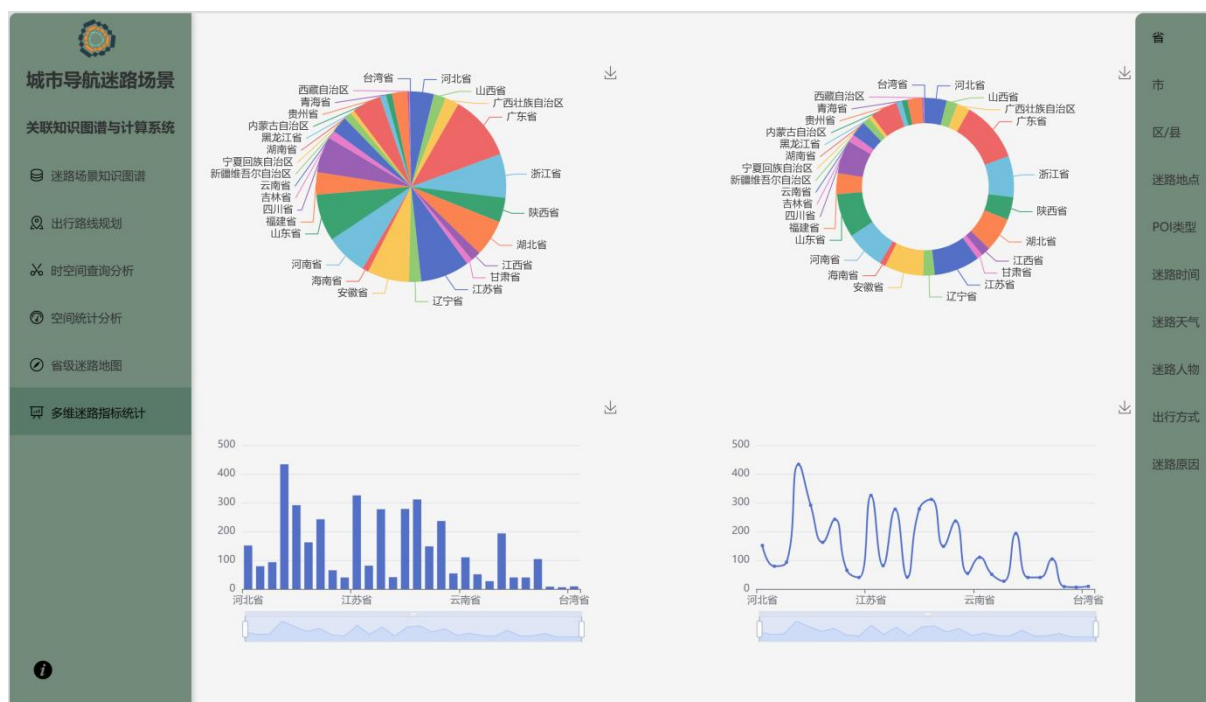




图 10. 多维迷路指标统计图

## 四、组织设计

### 1. 团队分工

团队分工的合理安排是项目成功的关键之一。我们的团队由专业的前端、后端和算法工程师组成,每个成员都有自己的专业领域和技能优势。我们将充分利用团队成员的专业能力,实现导航知识获取和决策推荐系统的高质量开发。

1) 数据分析能力: 我们的团队成员都具有良好的数据分析能力,可以对导航经验信息



进行有效的数据处理和分析，从而得出有价值的结论。

2) 人工智能技术：我们的团队成员对人工智能技术背景有深入了解，可以使用自然语言处理、机器学习等技术进行情感分析、因果关系分析等任务。

3) 导航软件开发经验：我们的团队成员中有多名进行过合作软件开发经验的工程师，可以为项目提供技术支持和指导。

在前期准备方面，我们已经对导航经验信息进行了收集和整理，同时也搜集了相关的数据分析和人工智能技术，为项目的顺利实施奠定了基础。因此，我们有信心能够顺利完成本项目，并为导航软件的改进做出贡献。

成员分工如下：

李安南：前端界面设计与实现，后端接口整合与测试

李思琪：NLP、知识图谱相关的算法设计、开发与测试

李梦萍：后端接口实现，数据库设计与概念模型设计

方天桐：前端界面设计与开发，前后端数据请求转发

## 2. 对知识和工具的掌握

### 1) 数据收集和处理

在导航知识获取和决策推荐系统的开发过程中，我们需要获取大量的群集导航经验文本数据集。为了保证数据的质量和一致性，我们需要熟悉数据爬取技术和数据清洗和预处理的方法。我们将使用 Python 语言和相关的数据处理工具，例如 pandas、numpy、scipy 等，对数据进行清洗、处理和分析。

### 2) 自然语言处理 (NLP)

自然语言处理是导航知识获取和决策推荐系统中不可或缺的技术。我们需要了解和掌握自然语言处理的基本概念和技术，例如词嵌入、词袋模型、文本分类、文本生成等。在这个项目中，我们将使用 Python 和相关的 NLP 工具库，例如 nltk、spacy 等，对导航经验文本数据进行处理和分析。

### 3) 深度学习模型

深度学习模型是目前处理自然语言处理和推荐系统等领域的最先进技术之一。我们需要学习和掌握深度学习模型的基本原理和常用算法，例如循环神经网络 (RNN)、长短期记忆网络 (LSTM)、卷积神经网络 (CNN) 等。我们将使用 Python 和相关的深度学习框架，例如 TensorFlow、PyTorch 等，设计和实现深度学习模型，以提取导航经验文本数据中的有用信息。

### 4) 知识图谱构建和查询

知识图谱是导航知识获取和决策推荐系统中的重要组成部分。我们将了解知识图谱的构建方法，包括实体抽取、关系抽取和图谱构建等技术。我们将使用 Python 和相关的图谱构建工具，例如 Neo4j、RDFlib 等，构建和查询知识图谱。

### 5) 机器学习和推荐系统

机器学习和推荐系统是导航知识获取和决策推荐系统中的另外两个重要技术。我们将了解机器学习的基本概念和算法，例如决策树、支持向量机、随机森林等，以及推荐系统的原理和方法，例如基于内容的推荐、基于用户的推荐、基于协同过滤的推荐等。我们将使用 Python 和相关的机器学习和推荐系统工具，例如 scikit-learn、surprise 等，设计和实现个性化的导航知识获取和决策推荐系统。

### 6) 编程和数据分析能力

编程和数据分析能力是导航知识获取和决策推荐系统开发过程中不可或缺的能力。我们将使用 Python 编程语言进行实现和模型训练，并使用相关的编程工具 and 数据分析工具，例

如 Jupyter Notebook、PyCharm、Git 等，进行代码管理和数据分析。

## 五、项目实施方案

### 1. 具体实施内容

从经验文本数据中产生与空间认知协同的经验知识语义表征与量化计算方法，实现导航经验语义要素化和特征化以后的可共享表达，为应用提供技术支撑。从目前经验素材的表达形式上，群集经验主要以基于社交媒体文本形式存在，依据导航关联语义框架中归置与划分的经验语义类别，利用深度学习技术，分类进行经验语义协同计算。

#### 1) 原始迷路导航经验信息管理

原始导航经验信息管理是指对用户出行过程中的导航经验信息进行收集、整合、存储、管理的过程。收集用户在使用导航软件过程中的一些基本信息，如用户姓名、出行目的、出行起点和终点等，将用户输入的信息整合起来，形成一条完整的导航经验信息。将整合好的导航经验信息存储到数据库中，以便后续进行数据分析和处理。

#### 2) 对导航迷路经验信息进行时空要素抽取并展示

对导航经验信息进行时空要素抽取并展示，是指从用户在出行过程中的导航经验信息中提取时间、天气、地点、环境因素等信息，并将其在地图上展示。利用实体抽取技术，从导航经验信息中提取时间、天气、地点、环境因素等信息，使用 BERT 模型进行模型训练和预测。通过导航软件提供的 API，获取用户信息，将提取到的时空要素进行处理，将时间、天气、地点、环境因素等信息进行分类和标注，便于后续展示和分析。处理好的时空要素信息在地图上展示。

#### 3) 对导航迷路的场景进行分类和统计展示

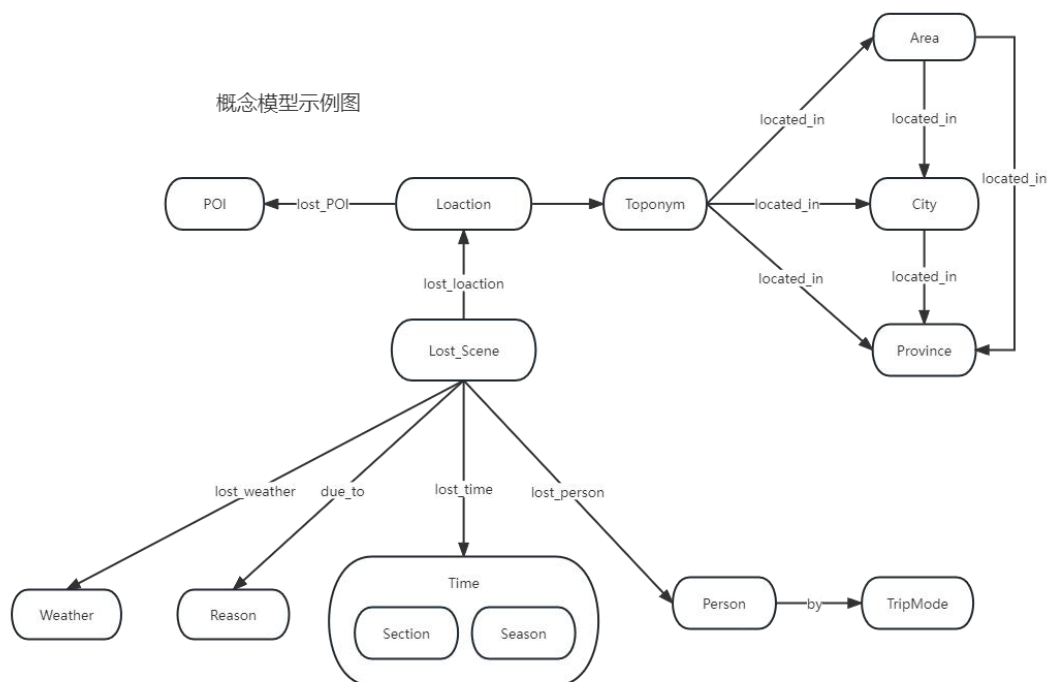
对导航迷路的场景进行分类和统计展示，是指将导航经验信息中用户迷路的场景进行分类和统计，以便用户更好地了解迷路的原因和场景，并规避这些场景，减少迷路的发生。对导航经验信息进行清洗和筛选，只保留用户迷路的信息，去除其他无用信息。对迷路场景进行分类，可以将迷路场景分为商场、胡同、立交桥等场景，将迷路场景知识图进行建模，并将其存储到 HugeGraph 分布式图数据库中，便于后续统计和分析，使用柱状图、饼图等形式进行展示，便于用户了解不同场景的迷路情况。

#### 4) 对导航迷路场景的知识图进行计算、展示与查询

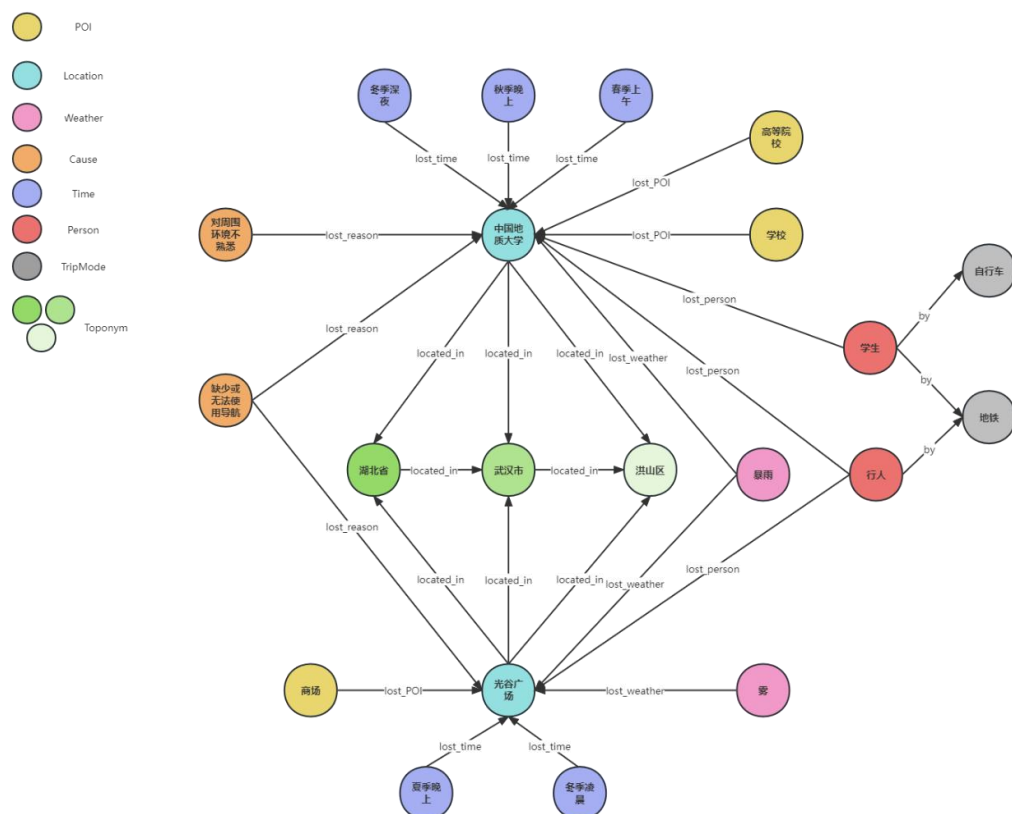
对导航迷路场景的知识图进行计算、展示与查询，将导航经验信息中用户迷路的场景进行抽象化建模并存储到 NEO4J 图数据库中，对知识图谱进行分析，并确定实体之间的关系和属性，并通过计算、展示和查询等方式，让用户更好地了解不同场景的迷路情况，为用户提供更好的导航体验。对导航经验信息进行清洗和筛选，去除其他无用信息，将不同场景的迷路情况构建成知识图谱进行展示。

### 2. 模型和系统设计

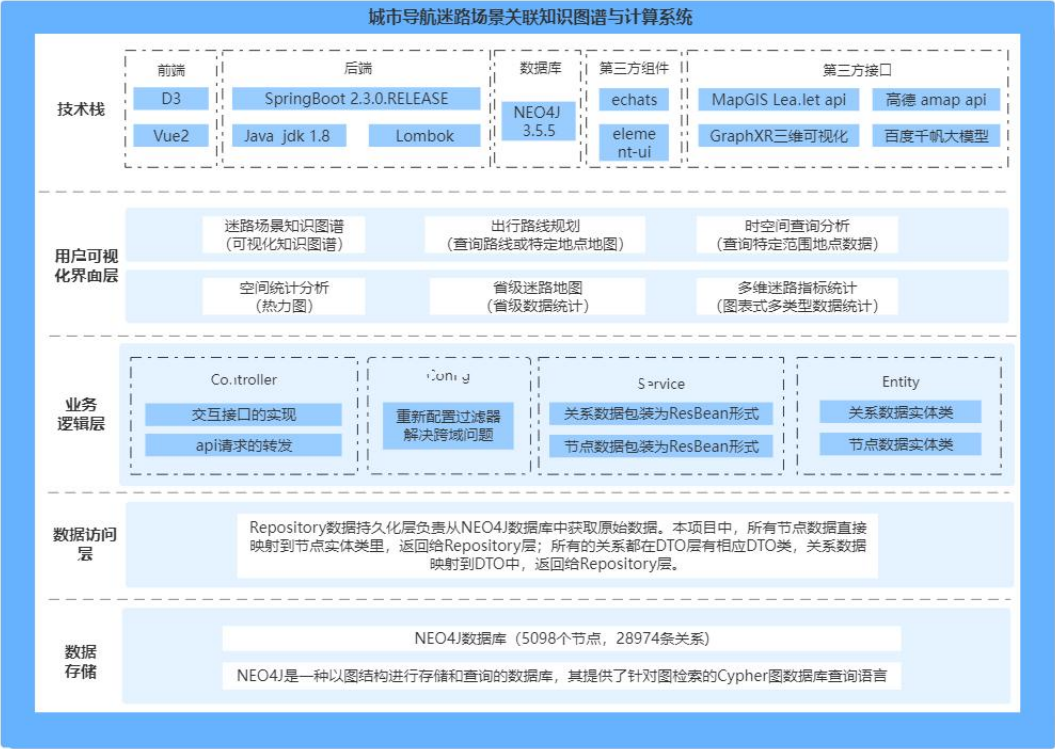
概念模型示例图



以具体迷路地点（Location）为中心节点，与该地点数据所处的区/县（Area）、市（City）、省（Province）、迷路人物（Person）、迷路时间（Time）、迷路原因（Reason）、迷路天气（Weather）、POI 属性（POI）直接相关联，构成数据库架构的基础要素；通过迷路时间（Time）、迷路原因（Reason）、迷路天气（Weather）、迷路人物（Person）与其交通方式（TripMode）相关联，以此构建迷路场景（Lost\_Scene）；区/县（Area）与市（City），区/县（Area）与省（Province），市（City）与省（Province）之间相互关联，构成行政区划（Toponym）要素。



如图可以发现，实例模型提供了以下导航经验：  
迷路地点、迷路时间、迷路原因、上级行政区划、天气、迷路当事人以及他们所乘坐的交通工具。  
通过实例图可以观察到，在冬季深夜、秋季晚上、春季上午等时间节点在中国地质大学容易发生迷路事件，该地点隶属于湖北省武汉市洪山区，迷路人员主要为行人和学生群体，在暴雨天气搭乘地铁或自行车时容易因为对周围环境不熟悉或缺乏导航发生迷路现象。



系统设计部分：在技术上，在本项目中我们前端主要用了 D3 与 Vue2 框架，后端使用 Spring Boot 框架，数据库为图数据库 neo4j。项目主要分为三个层次，分别是用户可视化界面层，业务逻辑层和数据访问层，三个层次使得本项目具有更好的解耦性。

图 11. 系统及数据库设计视图

### 3. 方案可行性分析

- 1) 数据来源：本研究所需的群集导航经验数据来自城市中的导航经验大数据，数据来源可靠。
- 2) 技术支持：本研究所需的技术支持包括自然语言处理技术、深度学习网络技术等，这些技术已经有较为成熟的应用案例，具备可行性。
- 3) 实验验证：本研究提出的方法需要进行实验验证，可以通过构建样本库、建立实验场景等方式进行验证。

### 4. 环境分析

- 1) 政策环境：政府大力推进智慧城市建设，城市导航服务是智慧城市建设的重要组成部分。政府出台的相关政策和规划，为城市导航迷路场景关联知识图谱与计算研究项目提供了政策支持和市场机遇。
- 2) 技术环境：人工智能技术、大数据技术和云计算技术等快速发展，为城市导航迷路场景关联知识图谱与计算研究项目提供了技术支持和保障。此外，移动互联网的普及和智



能手机的普及，为该项目提供了广泛的应用场景和用户基础。

3) 经济环境： 城市导航服务市场前景广阔，市场需求旺盛，可以为城市导航迷路场景关联知识图谱与计算研究项目提供广阔的市场空间和商业机会。此外，城市导航服务可以为旅游、交通、商业等多个领域提供服务，具有广泛的应用前景。

4) 社会环境： 城市导航服务已经成为人们出行和生活的重要组成部分，对于提高居民的生活质量和城市的服务水平具有重要意义。城市导航迷路场景关联知识图谱与计算研究项目的实施，可以为人们提供更为准确、全面的导航服务，提高城市的服务水平，满足人们出行和生活的需求，具有良好的社会效益。

综上所述，城市导航迷路场景关联知识图谱与计算研究项目具有良好的政策、技术、经济和社会环境，具有广阔的市场前景和应用前景。

## 六、 项目展望

### 1. 国内外研究现状

城市导航迷路场景关联知识图谱与计算研究是鲜有人关注的领域，目前与其相关性较高的研究领域是基于社交媒体的交通信息抽取研究，本研究大量参考了基于社交媒体的交通信息抽取领域中关于任务定义、抽取方法、评价指标等相关理论知识。基于社交媒体的交通信息抽取研究一直是交通研究领域的热门课题之一，许多研究人员提出了利用自然语言处理技术从社交媒体文本数据中提取交通信息的方法，这类方法的提出在一定程度上解决了人工提取交通信息耗时长、效率低等问题。其中，从社交媒体文本中提取交通信息的方法主要有文本分类、命名实体识别、无监督聚类等，本节将从文本分类、命名实体识别、无监督聚类三个 NLP 任务的研究现状引入到基于社交媒体的交通信息抽取的研究现状，并讨论目前交通信息抽取任务框架迁移到导航迷路经验建模与挖掘任务上的局限性。

#### 1) 城市知识图谱

如何将知识图谱建模，图数据库往往是一个优秀的选择。知识图谱是一个由实体、属性和它们之间的关系构成的巨大图形，而图数据库则是一种专门用于存储和查询这种大规模图形的数据库。由于其天然的图形结构，图数据库可以更好地处理实体之间的关系，比如，它们可以轻松地存储和查询实体之间的复杂关系。在建模知识图谱时，使用图数据库可以更好地捕捉实体之间的关系，并能够更加高效地查询这些关系。Yachen Tang 等人提出了一种基于图形数据库的知识图谱存储和查询方法，并在构建的电力设备管理知识图谱上比较了 RDF 通用存储格式和图形数据库的存储和查询性能。Pengcheng Liu 等人基于 Neo4j 图形数据库平台构建的台风知识图谱，可以将台风数据中的各种信息相互关联，进行全面和相关的分析，并最终为台风预防和灾后保护提供有效的帮助。Dongyang Zheng 等人同样利用了 Neo4j 图形数据库平台构建了学者-课程知识图谱（SCKG）。

#### 2) 文本分类

文本分类是自然语言处理(NLP)中的一个经典问题，旨在为句子、查询搜索字段、段落和文档等文本单元分配标签或标记。文本数据可以来自于电子邮件、社区聊天、社交媒体、用户评论等等网络平台。由于文本具有非结构化性质，所以文本分类会是一项具有挑战和耗时的工作。

文本分类有两种方式：（1）人工手动标注分类；（2）自动文本分类。

随着互联网的快速发展，文本的数据规模与日俱增，自动标注分类的方式才能适应数据规模的爆炸式增长。自动标注分类中又以监督学习为主，目前自动标注分类的方式主要有三种：（1）基于规则；（2）基于传统机器学习；（3）基于深度学习。

传统机器学习方法依赖于文本的向量化表示以及特征工程的预处理，深度学习方法为解

决手工特征的限制应运而生。深度学习方法的核心是使用自动学习的词嵌入模型，它将文本映射为一个低维连续特征向量，因此不需要手工生成特征。最早的词嵌入模型之一是 Dumais 在 1989 年提出的潜在语义分析 (LSA)，LSA 是一个在 20 万个单词上训练且不到 100 万个参数的线性模型。2001 年，Bengio 提出了第一个神经语言模型，该模型是一个基于 1400 万单词训练的前馈神经网络。然而，这些早期的词嵌入模型得到的词向量特征不如手工特征，因此没有被广泛使用。当使用大量的文本训练数据配合更大更复杂的词嵌入模型时，计算范式开始发生转变。2013 年，谷歌开发出了使用 60 亿个单词进行训练的 Word2VEC 模型，并迅速在许多 NLP 任务中流行起来。2017 年，AI2 和华盛顿大学的团队开发了一个基于 3 层双向 LSTM 的上下文词嵌入模型，该模型在 10 亿个单词上训练了 9300 万个参数，此模型被称为 Elmo。因为双向 LSTM 模型能够捕获文本的上下文语义，Elmo 比 Word2VEC 在有些任务上表现得更好。2018 年，谷歌开发出了一种新的神经网络架构——Transformer 来构建词嵌入模型，Transformer 是一类基于注意力机制的模型。同年，谷歌开发了基于双向 Transformer 的 BERT。BERT 由 3.4 亿个参数组成，使用 33 亿个单词进行训练，是当前最先进的词嵌入模型。目前使用更大模型和更多训练数据在自然语言处理深度学习中仍然是大势所趋。近期开发的模型印证了这点，OpenAI 的 GPT-3 模型包含 1700 亿个参数，Google 的 gshard 包含 6000 亿个参数。

尽管这些超大参数规模的深度学习模型在各种 NLP 任务上得到了优异的性能表现，但是仍然在许多任务上预测效果很不稳定，而且深度学习模型复杂的计算方式和超大规模的训练参数导致训练时间很长。目前，传统机器学习方法与深度学习模型在不同的任务场景下都有自己的席之地，没有绝对的全面优势，并且经典机器学习方法与深度学习模型的融合研究也越来越常见。

### 3) 实体抽取

命名实体识别 (NER) 是指从文本中识别并分类各种命名实体的自然语言处理任务。中文命名实体识别相较于其他语言更为复杂、困难，中文最明显的特点是词界模糊，没有定界符表示词界。而在英文中，词与词之间有分隔符来标识界限。另外，中文的语义比英文复杂，汉字能够灵活组合，语义表达丰富，在不同的上下文语境中表达不同意思。汉字作为中文文本的基本单位，模糊的词边界和复杂的上下文语义会造成大量的歧义，增加了中文命名实体识别的难度。

传统的中文命名实体识别方法主要有两类：基于规则的方法和基于统计的机器学习方法。基于规则的方法主要是使用正则表达式或字典，根据一些匹配规则从文本中检索实体。比如提取人名时，可以使用正则表达式采取“姓”加“名”的方式提取，而基于字典的方法则需要建立一个包含人名的词汇表。还比如，在提取地名实体时，基于字典的方法会建立一个地名录，这种方式至今仍是可靠的。

由于命名实体识别可以简单理解为实体分类任务，所以可以使用基于统计概率的方法做出预测。传统基于统计的机器学习模型包括隐马尔可夫模型 (HMM)、条件随机场 (CRF)、支持向量机 (SVM) 等，已经成功地用于命名实体的序列化和识别，并取得了良好的效果。

综上所述，基于社交媒体的交通信息抽取研究已经取得了许多具有启发意义的重要研究成果，然而交通信息抽取研究忽略了这种交通信息在时空上积累的导航迷路经验信息，亟待开展迷路场景知识图与计算研究，以更好地辅助城市导航服务，建立基于更广泛群集经验认知的经验共享与学习通道。

## 2. 项目特色及创新点

1) 建立导航迷路场景的知识图，将迷路场景的时间、空间、天气、地形、道路、建筑等环境特征等概念进行抽象与关系建模。该研究首次构建基于群集导航迷路经验的认知语义

框架,将泛在导航迷路经验文本转化为可学习的导航决策知识。通过构建城市导航知识图谱,将城市导航信息进行结构化和标准化,实现信息的智能化处理和推荐,提高导航信息的准确性和全面性,为用户提供更为优质的导航服务。

2) 面向导航迷路场景的知识图自动化查询计算方法,并针对迷路场景的特定查询构建图存储方案以及查询优化策略。利用人工智能技术,结合大数据分析和机器学习算法,实现导航信息的智能化处理和推荐,提供更为准确、全面的导航服务,提高导航信息的实时性和个性化,为用户提供更为便捷的出行体验。

3) 建立导航迷路场景的情感计算方法,基于深度学习网络建立导航要素的因果驱动与关联计算方法,对迷路场景的原因、环境、情感进行理解与计算。能够更加细致和深入地基于输入的文本分析用户情感,为用户提供更为体贴的导航服务。

4) 将导航迷路关联知识子图与地图要素进行融合与可视化呈现,增加用户的空间知识获取,强化其自主导航能力。通过地图等可视化方式,展示导航信息和路线规划,让用户更加清晰地了解导航情况和路线规划,提高导航信息的可视化程度,为用户提供更为直观的出行体验。

综上所述,本项目通过构建城市导航知识图谱,将城市导航信息进行结构化和标准化,实现信息的智能化处理和推荐。通过提高导航信息的准确性和全面性和友好易用的可视化用户界面,为用户提供更为优质、便捷、体贴和直观的迷路导航服务。