

Automatic Acquisition of Real-Time Traffic Data Based on Web Crawlers

Wenhao Yan, Yuqin Shu*, Zhiqin Huang

School of Geography, South China Normal University, Guangzhou Guangdong
Email: shuyq@m.scnu.edu.cn

Received: May 24th, 2016; accepted: Jun. 17th, 2016; published: Jun. 20th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Real-time traffic data is an essential data source to research fields such as intelligent city, low-carbon city, optimum of road net and so on. But the real-time traffic data is not available for free and public. In this paper, the Python programming language and the Tornado Net Framework are used to design a stable, efficient and timely web crawler program which could grab 1723 roads' traffic data of Guangzhou city, including its basic attributes and real-time traffic data every 5 minutes for each road, from four-dimensional traffic index page and save the data in the local MySQL database at the same time. The result shows that the web crawl technology is feasible and efficient in acquiring real-time traffic data.

Keywords

Web Crawler, Automatic Acquisition, Real-Time Traffic Data, Python Programming Language, Tornado Net Framework

基于网络爬虫的实时交通数据自动采集

闫文豪, 舒娱琴*, 黄植钦

华南师范大学地理科学学院, 广东 广州
Email: shuyq@m.scnu.edu.cn

收稿日期: 2016年5月24日; 录用日期: 2016年6月17日; 发布日期: 2016年6月20日

*通讯作者。

摘要

实时交通数据是低碳出行、智能交通、道路网络优化等研究领域必不可少的数据源。为了解决目前存在的实时交通数据不免费、不公开等问题,本文应用Python语言和Tornado网络框架,设计了一个稳定、高效、及时的爬虫程序。以广州市为例,从四维交通指数网页抓取到了1723条道路的基本信息及其每5分钟更新一次的实时交通数据,并将获取结果保存到MySQL数据库。结果表明网络爬虫技术在实时交通数据采集方面具有可行性和有效性。

关键词

网络爬虫, 自动采集, 实时交通数据, Python语言, Tornado网络框架

1. 引言

随着互联网技术的快速发展,网络信息资源也在飞速增长,越来越多的公司、企业和政府部门开始在自己的网站上开放自己的数据。有些数据是长期积累的历史数据,如统计年鉴数据;有些则是实时观测得到的定时更新数据,如气象、实时交通数据;有些数据可以直接下载,有些则限制了用户的下载权限[1]-[3]。虽然网络上提供的数据种类多样化,数据量也急剧增长,但网络上公布的数据大多只能浏览却不能直接下载。如何将丰富多样的数据下载下来,并把它们保存到本地数据库中,对以数据分析为基础的研究者来说是十分重要的[4]。实时交通数据为智能交通、低碳城市建设和城市环境保护研究提供了重要的数据基础,道路的实时交通数据如交通指数和平均车速是应用领域需要考虑的重要指标[5] [6]。政府和部分公司通过智能交通系统、浮动车等技术已经获取到了城市道路的实时交通数据,但考虑到诸多的问题却鲜有公开。四维图新公司作为三大路况信息提供商之一,将通过浮动车模型获取到的实时交通指数数据发布到了其官方的网站上,但是只允许浏览却不提供下载权限。

网络爬虫源于搜索引擎技术,主要任务是及时、高效地对互联网网页的搜集,并将收集到的网页保存到本地,提供给其它模块进行分析、挖掘和建立索引使用[7]。网络爬虫的类型主要可分为通用型网络爬虫、增量式爬虫和面向主题的爬虫,每种类型各有优缺点[8]-[11]。面向主题的爬虫,适合针对特定网页数据的获取,同时可根据网页的内容确定高效的抓取策略[12]-[14]。

本文采用面向主题的爬虫方法,以四维图新交通指数网站为对象,针对其公布的只读实时道路交通指数数据,使用Python语言和Tornado网络框架设计了一个网络爬虫程序。该程序对其网页进行分析,提取出数据请求的地址和参数,模拟浏览器请求服务器,抓取网站上发布出的广州市实时交通指数数据和道路的基本信息,并针对数据特点建立数据库,将抓取到的数据批量同步保存到本地数据库中,为数据分析及可视化做准备。

2. 实时交通数据

在进行爬虫程序设计前,对四维图新实时交通指数网页进行分析,确定网页上实时交通数据请求方式、请求参数及其相关的数据类型,然后根据交通数据及道路关系定义数据结构和建立表关系,构建一个交通数据库来存储爬取到的结果。

2.1. 四维交通指数网页

四维交通指数网页是四维图新公司发布的交通信息大数据产品之一。四维交通指数网页上以表格的

形式列出了实时交通数据，如图 1 所示，包括道路名称、道路起点、道路终点、交通指数、平均车速、道路等级和拥堵等级。交通指数是，结合道路实际速度及道路通行条件等的基础，加入对交通拥堵的主观感受程度，并用量化方法来表达道路的交通运行状况。通俗来讲，是指某条道路的实际测试，与该道路上驾驶员自由行驶速度的比值，比值越大，道路越拥堵[15]。四维交通指数网页将城市道路分为 1、2、3、4 四个等级，分别对应的道路类型高速公路、城市快速路、地面主干道和次干道。道路拥堵等级实际是符号化的交通指数值。列表中的每条记录可以实现属性到空间的查询，如图 2 所示。图 2 中红色标识的道路是根据经纬度数据重新绘制的。道路的坐标值数据和道路实时路况数据都以 JSON 格式存储。JSON(Java Script Object Notation)是一种轻量级的数据交换格式，包括对象和数组两种结构，通过这两种结构表示复杂的数据结构，对象为“{}”括起来的内容，数据结构为{key:value,...}，数组为“[]”括起来的内容，数据结构为[“java”,...] [16]。

2.2. 实时交通数据库

根据上面的网页信息，本文设计了道路基本信息表(见表 1)、道路坐标表(见表 2)和实时交通数据表(见表 3)分别存储道路基本数据、道路坐标数据和实时交通指数数据。

道路信息列表							
<input checked="" type="checkbox"/> 高速公路 <input checked="" type="checkbox"/> 城市快速路 <input checked="" type="checkbox"/> 地面主要道路			■ 畅通 [0-2) ■ 基本畅通 [2-4) ■ 缓慢 [4-7) ■ 拥堵 [7-10) ■ 严重拥堵 [10-18) ■ 路网瘫痪 (>=18)				
名称	起点	终点	交通指数	平均速度	道路等级	拥堵等级	
1 大广高速公路	京开高速公路	柏村	16.0	12.9	高速公路	■	
2 西直门北大街辅路	学院南路	西直门桥	9.2	18.2	地面主要道路	■	
3 建国门南大街	北京站东街	明城墙遗址公园	8.8	16.5	地面主要道路	■	
4 鸭子桥路	广安门南滨河路	广安门南街	8.7	16.5	地面主要道路	■	
5 京通快速公路	北苑南路	建国路	8.0	40.8	高速公路	■	
6 学院路	清华东路	西直门(西直门北大街)	7.8	33.3	城市快速路	■	
7 学院路辅路	北四环中路辅路	北土城西路	7.7	18.8	地面主要道路	■	
8 东三环	成寿寺路	三元桥	7.6	35.8	城市快速路	■	
9 莲花池西路	沙窝桥	莲花桥	7.5	32.0	城市快速路	■	
10 学院路辅路	北土城西路	北四环中路辅路	7.3	25.4	地面主要道路	■	

Figure 1. List of real-time road traffic situation
图 1. 实时路况信息列表图



Figure 2. Real-time traffic information of the Daguang Highway
图 2. 大广高速公路实时路况信息图

Table 1. Data structure of the roads' basic attributes
表 1. 道路基本信息表结构

编号	字段名称	字段说明	字段类型	备注
1	Id	编号	int	非空
2	RoadId	道路 ID	varchar(100)	主键
3	Roadname	道路名称	varchar(50)	非空
4	Length	道路长度	int	米
5	Startname	道路起点	varchar(50)	
6	Endname	道路终点	varchar(50)	
7	Dir	道路方向	char(8)	
8	Grade	道路级别	int	

Table 2. Data structure of the roads' coordinates
表 2. 道路坐标表结构

编号	字段名称	字段说明	字段类型	备注
1	Id	编号	int	非空
2	Lon	经度	double	度
3	Lat	纬度	double	度

Table 3. Data structure of the roads' real-time traffic data
表 3. 实时交通数据表结构

编号	字段名称	字段说明	字段类型	备注
1	Id	编号	int	非空
2	RoadId	道路 ID	varchar(100)	
3	Date	更新时间	char(16)	分钟
4	Cindex	交通指数	float(4)	
5	Avgspeed	平均车速	float(4)	km/h

为了便于数据的更新及后期管理，实时交通数据以天为单位，存储到一个表中，表名为当天日期，涉及字段包括：道路 ID、数据更新时间、交通指数、平均车速，数据结构如表 3 所示。

实时交通数据表和道路坐标表分别设置道路 ID 字段和道路名称字段为主键，并参考道路基本信息表中外键值，构成数据库中的表关系，如图 3 所示。

实时交通数据存储时，道路 ID、道路名称、道路起点、道路终点字段长度需要做特别的设置。四维交通指数网页中提供的道路名字较长，道路的 ID 字段值为道路名称的拼音组合，为了避免无法完整存储道路 ID 值，将道路 ID 字段属性设置为可变长的字符类型，最大长度设置为 100。MySQL 数据库因其体积小，速度快，方便管理和维护，并且免费，在中小型的应用中得到广泛使用[17]，因此本文将设计的表存储到 MySQL 数据库中。

3. 实时交通数据的自动采集

一个好的爬虫程序必须保证程序能够稳定、高效、及时地抓取到网页上的内容。一个完整的具有容

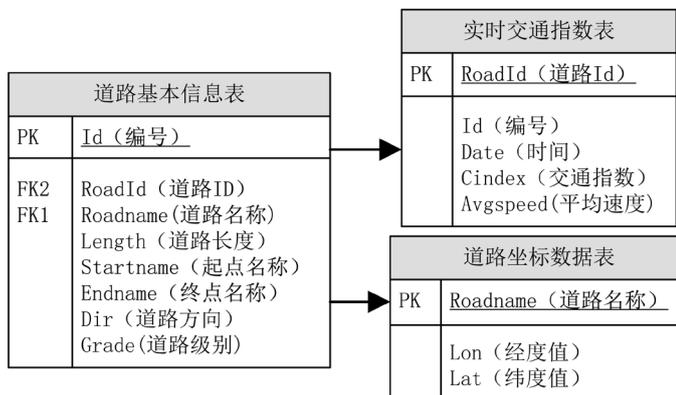


Figure 3. Relationships of tables in MySQL database

图 3. MySQL 数据库中表关系

错机制的处理流程是满足爬虫程序长时间稳定地运行的关键。在遇到不同的错误时，程序能够规避这些错误。在网上交互数据时，大部分的时间都花费在了服务的请求响应和数据的读写上面，特别是在处理大量可重复性的操作时，按照单条数据进行处理的时间花费是惊人的，实时交通数据读写同样面临着这样的难题。此外，网页数据一旦更新应该及时有效地进行抓取，即启动抓取程序，以获取实时交通数据。

3.1. 数据采集流程

本文采用主题型爬虫方式，针对特定的 URL 抓取数据，省去通用爬虫架构中的 URL 过滤、更新等模块，节省抓取流程时间，实现程序的高效率运行。要抓取的数据可分为实时交通数据和道路基本信息数据，实时交通数据每 5 分钟更新一次，道路基本信息数据固定不变，虽然这两者的更新策略不同，但是整个抓取数据的技术流程却是相同的。

数据抓取的技术流程是，首先输入特定 URL，以 POST 方式向服务器发送请求，这个过程中域名解析服务将域名解析为对应的 IP 地址，连接到服务器对应的端口，服务器接受到请求后，返回 http 应答包头信息，接着客户端判断服务器响应信息，如果服务器有响应，则客户端会根据返回的内容读取页面内容，判断读取到的数据是否为空，如果不为空，根据关键字段解析出实时交通数据，将交通数据按一定结构进行重组，最后连接数据库，将重组后的数据批量写入到数据库中。整个流程结束后，设置下次开始抓取的时间，程序按照上面的流程循环进行数据抓取。数据获取的流程图如图 4 所示。本文将实时交通数据的自动采集分成数据抓取、数据解析与重组和数据存储三个模块以及一个休眠控制。

3.2. 程序实现

本文使用 Python 开发语言，结合 Tornado 网络框架，将数据抓取、数据解析、重组和数据存储模块集成，最终将抓取到的实时交通数据批量存储到 MySQL 数据库中，实现了一个获取实时交通数据的稳定、高效、及时的网络爬虫程序。

3.2.1. 数据抓取模块

数据抓取模块是整个程序实现的关键，任务就抓取到网页上的数据。网页的内容一般来源于网页标签、JavaScript 代码和 Ajax 异步请求等方式[18]-[20]，其中最常用的是 Ajax 异步请求，而且这种方式得到数据最容易处理，本文使用的就是这种方式。Ajax 异步请求就是使用 POST 或 GET 的方式，将需要请求的地址和参数一起发送到服务器端，服务器接受请求后，将数据返回到客户端，客户端接受并返回的

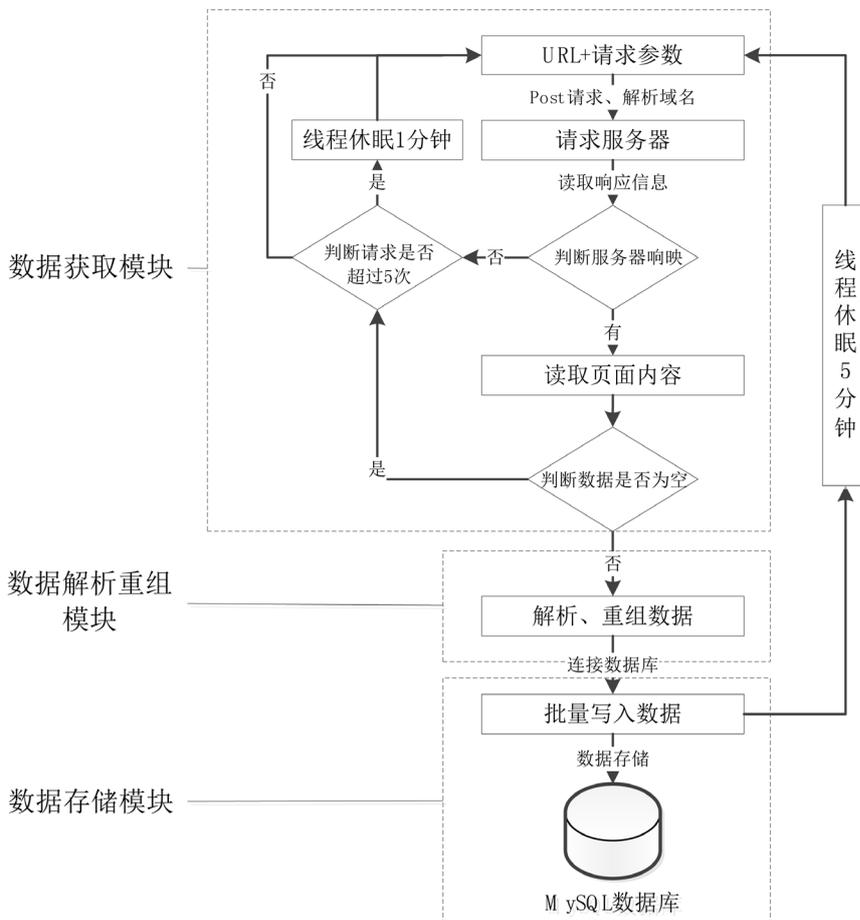


Figure 4. Workflow of the real-time traffic data fetching strategy
图 4. 实时交通数据获取的流程图

结构化或非结构化的数据，根据关键字段解析获取到的数据，按一定的结构将数据显示到网页上[16]。在 Tornado 网络框架下，模拟浏览器，使用网页分析得到的请求地址和参数，如城市编码、道路级别等，向服务器发送请求，服务器完成请求处理后，将处理数据返回，完成数据抓取。据四维网页分析可知，数据请求的地址为 <http://www.nittrafficindex.com/traffic/getRoadIndex.do>，实时交通数据请求的参数为广州市区域代码 4401000、道路级别(1, 2, 3, 4)和数据更新的时间 5 分钟，道路基本信息的请求参数为广州市区域代码和道路 ID。

整个数据抓取过程中需要注意很多问题。第一，为保证程序稳定性，使用 Tornado 网络框架，模拟客户端请求服务器，避免直接请求带来请求异常，同时避免服务器禁止特定 IP 对服务器的频繁访问。第二，网页上的交通数据每 5 分钟更新一次，一次数据获取结束 5 分钟后，再次启动抓取模块及时抓取更新后的交通数据。第三，抓取的过程中如果遇到数据服务器停止服务，导致响应结果异常，抓取不到数据，即在连续 5 次请求失败或返回的数据为空时，强制休眠程序，以避开不必要的硬件资源开销，在休眠一段时间后再重启抓取模块。第四，实时交通数据和道路基本信息数据，这两种数据请求使用相同的 IP 地址，但是实时交通数据使用循环抓取策略，道路的基本信息，无需重复抓取，一次获取就可得到。

3.2.2. 数据解析与重组模块

抓取模块结束后，获取的数据是 JSON 格式。根据 JSON 数据格式键值存储的特点，使用递归函数

将字段对应属性解析出来，再将数据按照对应的属性值顺序进行重组，保存到新建 List 数组中。爬取下来的实时交通数据和道路基本信息的属性有很多，包括时间、道路名称、道路 ID、平均车速、交通指数、道路级别、道路方向、道路起点、道路终点、道路长度、道路经纬度。解析后的道路经纬度信息是一条数据串，在实际使用的时候并不方便，需要重新进行解析，把坐标串分解为单独的坐标对，形成坐标点格式的数组，以方便后期道路可视化。

3.2.3. 数据存储模块

数据存储是把重组好的数据批量存储到本地 MySQL 数据库对应的表中。在数据存储时要考虑两方面，一个是实时交通数据的存储，另一个是道路属性数据和道路坐标数据的存储。前者是交通指数网页实时交通数据更新后，抓取程序获取到新的重组的实时交通数据后进行批量存储；后者是程序将抓取得到重组后的道路基本数据，单次批量写入 MySQL 数据库中。批量操作的优点使得单次流程的处理时间大大缩短。

3.2.4. 休眠控制

整个程序处理一个时间段的数据时，从数据抓取到数据解析，再到重组数据、批量插入数据库，平均耗时为 6 秒钟，如果不使用 Tornado 网络框架和数据批量写入的方法，一个抓取流程的耗时为 30 秒左右。程序在进行抓取数据时，根据实时交通数据的更新时间，每 5 分钟更新一次，计算出下次数据更新的时间，设置线程睡眠的时间，实现控制程序及时抓取到更新的交通数据，保证数据抓取延迟控制在数据更新后的 1 分钟以内。最终程序实现了稳定、高效、及时地获取道路实时交通数据。

4. 爬虫结果分析

本文应用 Python 语言和 Tornado 网络框架，设计了一个基于网络爬虫的实时交通数据自动采集程序。结果表明程序抓取到了广州市 1723 条道路的基本信息，每 5 分钟更新一次的道路交通数据。表 4 所示的为获取的道路基本信息，包括道路的名称、起点、终点、长度、方向等。表 5 为大学城环中路的经纬度坐标值。表 6 为 2015 年 10 月 22 日间隔为 5 分钟的道路交通指数数据及平均车速数据。由于每张表中包含的记录较多，这里只列出了部分数据。

图 5 为广州市 2016 年 3 月 1 日到 3 月 31 日四个级别道路全天的平均车速，图 6 为广州市 2016 年 3 月 1 日到 3 月 32 日四个级别道路全天的平均交通指数。

从表 5 中选取广州大学城的道路坐标值进行了矢量化处理，如图 7 所示。从表 6 中筛选了广州大学城环东路的实时交通指数数据和平均车速进行了图形化，见图 8。

5. 结束语

本文使用 Python 语言和 Tornado 网络框架，设计了一个稳定、高效、及时的网络爬虫程序。以广州市为例，从四维交通指数网页抓取到了 1723 条道路的基本信息及其每 5 分钟更新一次的实时交通数据，并将获取结果保存到 MySQL 数据库，实现了道路实时交通数据的自动采集，解决了实时交通数据获取难的问题。实验结果表明网络爬虫技术在交通数据采集方面具有可行性及有效性。网络爬虫技术不仅丰富了我们的获取数据的方法，而且充分利用丰富的网络资源，节省了研究者进行道路交通数据采集的成本和时间。

本文获取到的数据质量还需要进一步的验证，以期应用到低碳出行、智能交通、道路优化等研究领域。此外作为一个原型程序，还有很多的细节需要处理，如重复实时交通数据的识别和删除，可在数据写入数据库前进行处理，保证获取数据的质量。

Table 4. Basic attributes of the roads
表 4. 道路基本数据

ID	Road ID	RoadName	Startname	Endname	Length	Dir
1	4401000da4xue2cheng2nan2san1lu4	大学城南三路	大学城外环西路	大学城内环西路	1883	南
2	4401001da4xue2cheng2zhong1ba1lu4	大学城中八路	大学城内环东路	大学城外环东路	1927	东南
3	4401000da4xue2cheng2zhong1ba1lu4	大学城中八路	大学城外环东路	大学城内环东路	1638	东南
4	4401000da4xue2cheng2dong1er4lu4	大学城东二路	大学城外环东路	大学城内环东路	1762	东
5	4401001da4xue2cheng2xing1guang1xia4dao4	大学城星光下道	南沙港快速路	大学城外环西路	1759	南
6	4401001da4xue2cheng2bei3san1lu4	大学城北三路	大学城内环东路	大学城外环东路	1620	南
7	4401001da4xue2cheng2zhong1huan2dong1lu4	大学城中环东路	大学城中环西路	大学城中心大街南	8792	东北
8	4401000da4xue2cheng2zhong1huan2xi1lu4	大学城中环西路	大学城中环东路	大学城中心大街南	9269	东北
9	4401001da4xue2cheng2wai4huan2xi1lu4	大学城外环西路	大学城外环东路	大学城中三路	14925	西南
...

Table 5. Coordinates of the roads
表 5. 道路坐标数据

ID	Lon	Lat
1	113.39701	23.07153
2	113.39686	23.0713
3	113.39682	23.07122
4	113.39642	23.07059
5	113.39634	23.07047
6	113.39611	23.0701
7	113.39611	23.0701
8	113.39536	23.06886
9	113.39536	23.06886
...

Table 6. Traffic index and average speed of the 1723 roads in every 5 minutes on October 22nd, 2015
表 6. 2015 年 10 月 22 日间隔为 5 分钟的 1723 条道路交通指数及平均速度

ID	RoadId	Date	Cindex	Avgspeed
1	4401000da4xue2cheng2nan2san1lu4	0:00	3.3	26.1
2	4401001da4xue2cheng2zhong1ba1lu4	0:00	1.2	38.8
3	4401000da4xue2cheng2zhong1ba1lu4	0:00	0.1	40.5
...
1724	4401001da4xue2cheng2zhong1ba1lu4	0:05	2.6	29.3
1725	4401000da4xue2cheng2zhong1ba1lu4	0:05	1.2	34.3
1726	4401000da4xue2cheng2nan2san1lu4	0:05	0.6	47.1
...
144733	4401000da4xue2cheng2zhong1ba1lu4	7:00	2.4	29.3
144734	4401001da4xue2cheng2zhong1ba1lu4	7:00	0.6	37.7
144735	4401000da4xue2cheng2nan2san1lu4	7:00	0.6	37.1
...

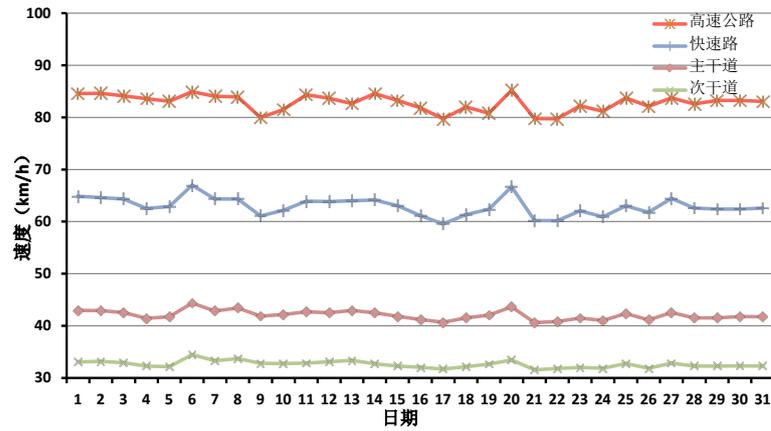


Figure 5. Daily average speed of different level roads in Guangzhou from Mar. 1 to Mar. 31, 2016

图 5. 广州市 2016 年 3 月不同级别道路的全天平均车速

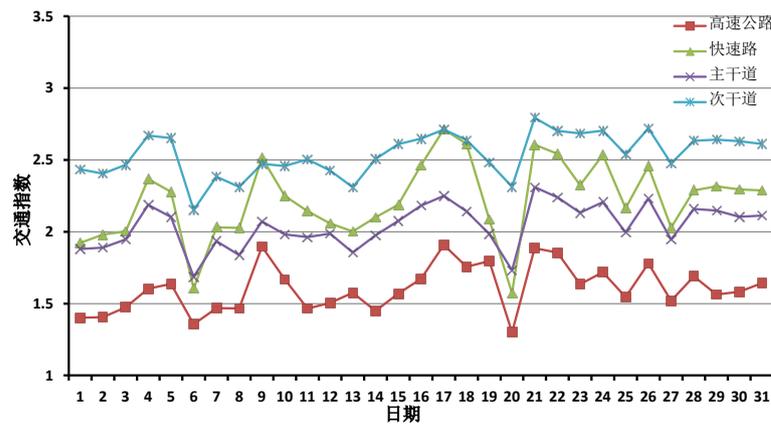


Figure 6. Daily average traffic index of different level roads in Guangzhou from Mar. 1 to Mar. 31, 2016

图 6. 广州市 2016 年 3 月不同级别道路的全天平均交通指数

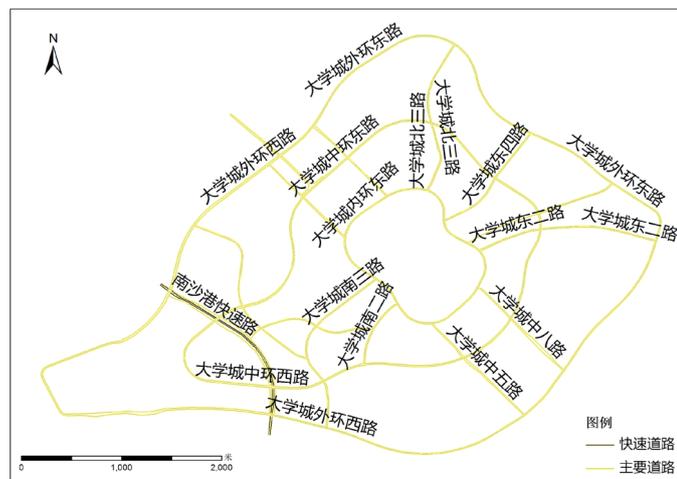


Figure 7. Visualization of the road coordinates in the Guangzhou Higher Education Mega Center

图 7. 广州市大学城道路坐标数据可视化

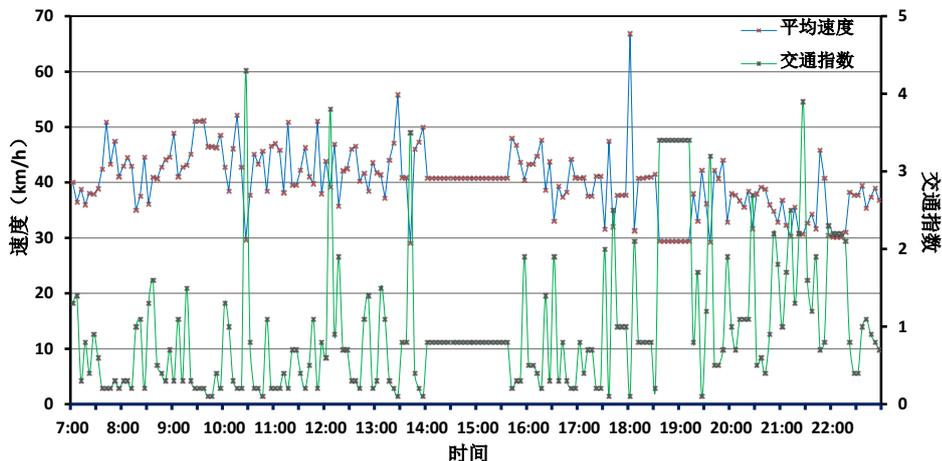


Figure 8. The traffic index and average speed of the East Daxuecheng inner ring road from 7:00 to 23:00 on October 22nd, 2015

图 8. 大学城环东路 2015 年 10 月 22 日 7:00~23:00 每 5 分钟更新一次的交通指数和平均车速

基金项目

国家自然科学基金项目(41101184), 教育部第 45 批留学归国人员科研启动基金项目(一种基于自底向上和自顶向下的城市交通碳排放的混合空间量化方法)。

参考文献 (References)

- [1] 于娟, 刘强. 主题网络爬虫研究综述[J]. 计算机工程与科学, 2015, 37(2): 231-237.
- [2] Kausar, M.A., Dhaka, V.S. and Singh, S.K. (2013) Web Crawler: A Review. *International Journal of Computer Applications*, **63**, 31-36. <http://dx.doi.org/10.5120/10440-5125>
- [3] 曹磊. 网络空间的数据权研究[J]. 国际观察, 2013(1): 53-58.
- [4] 刘兵, 等. Web 数据挖掘[M]. 北京: 清华大学出版社, 2013.
- [5] 陈美. 大数据在公共交通中的应用[J]. 图书与情报, 2012, 148(6): 22-28.
- [6] 周春梅. 大数据在智能交通中的应用与发展[J]. 中国安防, 2014(6): 33-36.
- [7] 李欢. 基于 API 天气数据抓取的特定网络爬虫的研究与实现[D]: [硕士学位论文]. 秦皇岛: 燕山大学, 2014.
- [8] 张春菊, 等. 基于网络爬虫的地名数据库维护方法[J]. 地球信息科学学报, 2011, 13(4): 492-499.
- [9] Li, W.W., Yang, C.W. and Yang, C.J. (2010) An Active Crawler for Discovering Geospatial Web Services and Their Distribution Pattern—A Case Study of OGC Web Map Service. *International Journal of Geographical Information Science*, **24**, 1127-1147. <http://dx.doi.org/10.1080/13658810903514172>
- [10] 陈晓慧, 陈荣国, 卫文学. 基于网络爬虫的 Web 服务抓取解析器的设计与实现[J]. 地理信息世界, 2010, 8(3): 64-68.
- [11] Palkowsky, B. and MetaCarta, I. (2005) A New Approach to Information Discovery—Geography Really Does Matter. *Proceedings of the SPE Annual Technical Conference and Exhibition*, Dallas, Texas, 9-12 October 2005.
- [12] 张红云. 基于页面分析的主题网络爬虫的研究[D]: [硕士学位论文]. 武汉: 武汉理工大学, 2010.
- [13] 郝以珍. 基于页面分析的网络爬虫系统的设计与实现[D]: [硕士学位论文]. 武汉: 华中科技大学, 2012.
- [14] 宋海洋, 刘晓然, 钱海俊. 一种新的主题网络爬虫爬行策略[J]. 计算机应用与软件, 2011, 28(11): 264-267.
- [15] 关积珍. 城市交通综合指数、交通出行指数及其数学建模[J]. 交通运输系统工程与信息, 2004, 4(1): 49-53.
- [16] 屈展, 李婵. JSON 在 Ajax 数据交换中应用研究[J]. 西安石油大学学报: 自然科学版, 2011, 26(1): 95-98.
- [17] 郑岚. Python 访问 MySQL 数据库[J]. 电脑编程技巧与维护, 2010(6): 59-61.

-
- [18] Faheem, M. and Senellart, P. (2013) Intelligent and Adaptive Crawling of Web Applications for Web Archiving. *Web Engineering*. Springer Berlin Heidelberg, 306-322. http://dx.doi.org/10.1007/978-3-642-39200-9_26
- [19] 杨波. 链接分析中的数据采集技术研究[J]. *图书馆理论与实践*, 2010(10): 44-48.
- [20] Yu, J. and Liu, Q. (2015) Survey on Topic-Focused Crawlers. *Computer Engineering and Science*, **37**, 231-237.

再次投稿您将享受以下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>