National Highway Toll Fare Estimation Using Polygon Bound Through Web Scraping

[1] *Redrouthu. Jahnavi,* [2] *Jetti. Gayathri Lakshmi,* [3] *Chintada. Hemanth Sai,*
[4] *Dwarampudi. Sri Rama Krishna Reddy,* [5] *Dwarampudi. Vijaya Sree Sai Reddy,* [6] *Jonnalagadda. Surya Kiran*

[1,2,3,4,5,6]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India -522302.

[1] *jahnaviredrouthu479@gmail.com,* [2] *gayathrilakshmijetti@gmail.com,*
[3] *hemanthchintada@gmail.com,* [4] *dsrkreddy81@gmail.com,* [5] *dwarampudisaireddy1514@gmail.com*
[6] *kiransurya93@kluniversity.in*

Abstract: In general, Toll Plazas are certan stands that take toll payments for continuing road maintenance and improvement. There are several ways to pay taxes, including cash, credit/debit cards, mobile wallets, smart cards, and RFID Electronic Toll Collection (ETC)/FASTAGs. To pay the tax via the aforementioned methods, a sufficient amount in the bank account is mandatory. You will be fined if you don't pay the toll tax at that time. Therefore, we have discovered a way to assist the user by notifying the toll fee information to them by providing the source and destination. The entire toll cost information for the journey has been obtained using an effective and straightforward technique that has been designed and implemented in this paper.

Keywords: Web Scrap, Toll Plaza, Toll Cost, Optimal Path, FASTag, Distance

I. INTRODUCTION

There is a large amount of data generated day to day. If we want to access a particular data from a large website, we cannot do it manually in a shorter period. Web scraping is used to simplify the procedure. Extracted data from a webpage is known as web scraping.There are many structured formats in which the scraped data can be stored.Scraping a website makes human work easier and faster. It gets the data required from the website very faster and in an easier manner.

Traveling on highways (toll roads) requires a toll tax to be paid by us. FASTag works using RFID (Radio Frequency Identification) technology which is fixed on vehicle windshields and makes payments directly from the bank/wallet. FASTag makes the work faster. This reduces the time to wait at the Tollgate to pay the toll tax.

Getting an idea of the total tax to be paid during the whole journey makes our trip economical. Also, we can make our payment at the toll much easier. According to our research, toll cost is calculated according to the vehicle type, source city, and destination city. We can also add a halt position where we can also go through that hat position so that we can get toll cost passing through that halt city. When we provide a source and a destination city, we search for the least possible way to pass through an intermediate city (If selected) and then we search for the toll gates available on that path. We also extract toll costs according to the vehicle type and the cost of each toll gate passing through that path.This gives us the total toll cost of the journey from the source and destination.

In getting the details of the toll cost, we used web scraping. We scraped a website that gives all the details of the toll gates available in India. It scrapes the toll cost and the details of the toll available on the website. First, it scraps the toll id from the list of tolls and loops over each website with the id as a parameter. This result will be stored in a file, mostly CSV, and used in calculating the toll cost.

This toll cost scraping is done in a periodical manner or in an emergency. As the government may increase the toll costs, we can scrap the data periodically so that the data with us is up to date. In maximum cases the changes may be effective from the first of a month, so we periodically extract every 1st of the month. If the government makes these changes effective from a particular date which is not the first of the month, we can start scraping the data manually.

This paper explains more about how the toll cost is calculated and how we researched how to scrap the data. Each section explains each module of our research. This includes,Web scraping, Toll cost calculation, application development, issues faced, limitations, and conclusions.

II. LITERATURESURVEY

A.Title: Web Scraping: Applications and Scraping Tools

Authors: Priya Matta,Devyani Sharma,Sachin Sharma, Nikita Sharma, Bhasker Pant

Every day, huge quantities of data beinggenerated. However, we cannot do it manually in a lower span of time, if we want to pierce a particular data from a large website. To make the process easier, web scraping is introduced. So, this paper discusses about different web scraping tools and operations. The scraped data can be stored in different structured formats. Web scraping is substantially used in exploration, business and educational purpose. There are different tools to scrap the data. The main drawback is occasionally web scrapers can get blocked by websites.[1]

B. Title: An Introduction to FASTag: A game changer in Automatic Toll collection systems in India

Authors: S.Amrin

This paper discusses about why FASTag is introduced in India. Travelling on highways (toll roads) requires a toll tax to be paid by us. To pay the toll tax FASTag was introduced in 2014 and made mandatory in 2021.Before the FASTag was introduced we usually pay the toll tax by waiting in queue and paying by cash which leads to waste of time, fuel waste, air pollution, and many more. To overcome all these problems FASTag was introduced. FASTag works using RFID technology which was fixed on vehicle wind shield and make payments directly from bank/wallet. Main drawbacks are FASTag charges are high for multi-axle vehicles, can be stolen, wrongly charged, may not be approved sometimes for payment. There is no specific website for FASTag service users. NHAI is not involving in FASTag payments itself.[2]

C. Title: Usage of the A* Algorithm to Find the Shortest Path in Transportation Systems

Authors:Akif Durdu, Esat Dere

This paper discusses, how shortest route in transportation is determined using an A* algorithm. On a Google map, the A* algorithm may be used to determine the shortest route between sources and destinations. In this algorithm the map is segmented as grid cells and based on the traffic density provided bygoogle maps, it will find the shortest path. The processing of the map portion on the algorithm map and the computation of the shortest path determined by the algorithm on the map are done manually, which is the major drawback. By creating the appropriate software and applying image processing methods, this operation may be carried out digitally. [3]

D. Title: A flow model for optimization of highway toll stations

Authors: Yifan Chen

This paper discusses about at which lane the vehicle can go depends on the traffic condition. Highway toll stations route optimization model deals with at which lane the vehicle must travel depends on the traffic density on the highway. If there is a congested toll booth, at which lane you can go is discussed by this model. When it comes to electronic toll collection (ETC), it's a free flow area, you can go in any lane. The main drawback is more extensions of the HITSROM model are not explained. [4]

E. Title: Analysis of Efficiency of Toll Road Network Development

Authors:M O Karpushko,E N Karpushko, V I Bryzgalov, I L Bartolomei, , A V Zhidelev

This paper mainly discusses about how efficiently toll road network is built in Russia. Toll prices in Russia is based on the vehicle speed it goes on a highwayTransponder (transmitting receiving device) is fixed to windshield of the vehicle. To reduce the cost of tolls, a free flow system is implemented, in which you must pay to travel at high speeds without encountering traffic jams on the road.The drawback in this paper is only about Russian toll networking system and its working.[5]

F. Title: A Heuristic Solution of the Toll Optimal Problem with Congestion Affected Costs

Authors:Vyacheslav Kalashnikov, Nataliya Kalashnykova, José G. Florez-Muniz, Vladik Kreinovich

This paper discusses about heuristic solution for setting of toll costs. Setting proper costs for toll roads while driving through a transportation network along paid arcs. The main drawback is to solve more complicated these problems, which includes several objective functions, the used technique has to be expanded.[6]

G. Title: A new path planning method based on concave polygon convex decomposition and artificial bee colony algorithm

*Authors:Zhaoying Li, Hao Liu, Zhao Zhang, Liang Yang*This paper discusses path planning using the artificial bee colony method and concave polygon convex decomposition. Offering an improvised free space technique that is based on map decomposition and uses an artificial bee colony algorithm along with several linked domains. In essence, a novel path planning approach is offered with the aid of the artificial bee colony algorithm and concave polygon convex decomposition. The biggest disadvantage is that during polygonising the map, some irregular maps may lose accuracy.[7]

H. Title: Estimating minimum toll rates in public private partnerships

Authors: Nevena Vajdic, Goran Mladenovic, Ceaser Queiroz

This paper discusses about estimating the toll costs when public, private partnership in tool booth. To look into the relation between toll rates level and technical project parameters along with financial parameters. For several input values a financial model is used to calculate the toll rates. This strategy is used to examine the selected parameters sensitivity to prospective changes on input values through the calculation of elasticities. The main drawback is change in operational costs and project construction the toll rates may vary.[8]

I.Title: Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application

Authors:Moaiad Ahmad Khder

This paper discusses how web scraping is used to acquire data that enables us to obtain pricing from retail locations very instantly and provides further information. As part of the working process of web scraping, applications and technologies like decision support system, knowledge engineering, data science, big data analytics, cyber security, as well as how we can relate to Python Language, are discussed. Additionally, a focus on highlighting the ethical and legal issues is given. The main drawback is that web scraping is currently a widely used and extremely practical technology in the information age. As a result of its quick learning curve and robust capabilities, Python is one of the finest languages for implementation.[9]

J. Title: A Combined Model of Congestion Toll Pricing Based on System Optimization with Minimum Toll

Authors:ShahriarAfandizadeh Zargari, HamidMirzahossein

This paper analyses how traffic management systems traditionally address the issue of congestion toll charging on transportation networks. the first is to lessen or avoid the obstruction, next is to secure financing for the development and upkeep of transportation improvements. Finding toll price based on system optimization with the lowest toll cost is the last step. principal flaws include While the flow distribution in network connections are pretty near to the system balanced conditions, Hern structure suggests that the number of chosen links for toll assignment is decreased in the benchmark urban transportation network (utnb). The direction of the connections' flow is not regarded as a fixed limitation. The network's equilibrium requirement can basically be fulfilled by increased entire flow, a lesser cost, and a roughly closer distance between the total price and the equilibration state.[10]

III.METHODOLOGY

Web framework consists of 3 modules
  i.   Web Scraping
  ii.   Admin
  iii.   Client/ End User

A. Module 1: In **Web Scraping,** we extract data from NHAI website and store into MySQL Database.

B. Module 2: **Admin** module is responsible for scraping data from NHAI website and to make sure that data is being updated periodically and getting stored into the Database.

C. Module 3: **End User** has the privilege to enter the source and destination of his travel where he gets to access all the toll information between that path including the total toll amount required for that path.

Algorithm 1: Web Scraping

Step 1: Invoking the web driver of chrome by using Selenium library in python.
Step 2: Load the required website using that web driver.
Step 3. Extract the html code to string format from the website by using the Beautiful Soup.
Step 4. Extract the required data like tollgate name, location, cost etc by using some string operation.
Step 5. Finally, store the extracted data in a required format such as csv, json etc

Algorithm 2: Polygon Bound & Toll Cost Estimation

Step 1: Collecting way points from source to destination by using Google Maps API.

Step 2: Pointing Upper bounds and lower bounds with respective to the path called as polygon area.

Step 3: Checking each tollgates location whether it is present at the polygon area or not.

Step 4: If it is present on the polygon area, then we can call that tollgate is located on the path and displayed on Map.

Step 5: After displaying each toll, then itself we are storing the toll cost into a variable.

Step 6: By adding all toll costs at each toll to the variable, finally we will get total toll cost between source and destination.

IV. BLOCK DIAGRAM
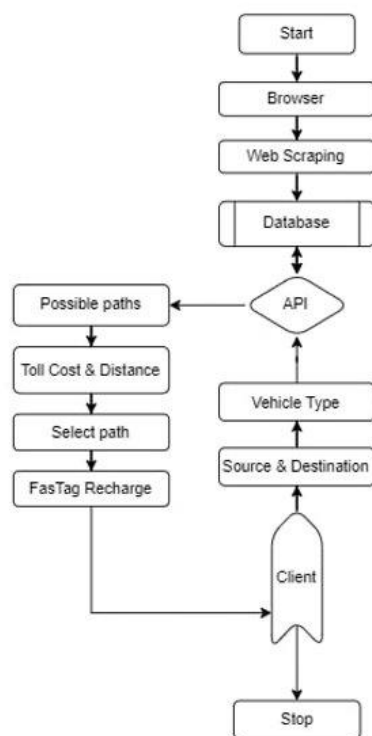
Fig 1:Workflow of Block Diagram

Step 1: Start

Step 2: Scrap the data from the browser and store it into database as a periodical update.

Step 3: Client will enter the source and destination with vehicle type in Web application.

Step 4: Calculate and display the Toll Cost & Distance of all possible paths using Google API.

Step 5: Select the optimal path to get the minimum recharge into the FASTag from client.

Step 6: Stop

V.IMPLEMENTATION

A. Web Application

1. HTML/CSS/JS

a.Hyper Text Markup Language (HTML): It is a fundamental method of expressing content and information on the web in a realistic manner.

b. Cascading style sheets (CSS): It is a language used for outlining the layout of the document that is formatted

with the help of markup languages like HTML/XML.

c. JavaScript (JS): It is a prearranging language that empowers you to make progressively refreshing substance, vitalize pictures, control multimedia and essentially all the other things.

2. Express JS:

6

4

This back-end web application framework is used to build RESTful APIs with the help of Node.js, basically it was a FOSS (Free and Open-Source Software) that is under the software license named MIT license. Since it is a web application design, it is mainly used for building web applications and APIs.

3. Node JS:

It is considered to be a server environment which is framed as an open-source environment. It runs on Windows, Linux, MacOS and so on. So, it is considered to be a cross platform. It runs on a JavaScript Engine and the JavaScript code is executed outside a web browser.

4. google-polyline library:

Google's Encoded Polyline Algorithm Format is implemented in Python as polyline (http://goo.gl/PvXf8Y).

It is simply a port of https://github.com/mapbox/polyline made with support for Python 2 and 3. A Google Maps Encoding Package polyline implements polyline encoder and decoder.

B. Admin Application

1. Django:

Django is a web application framework that is developed in Python. It is open source and cost-free framework to build the web pages in a simpler manner. It is more secure, fully featured server-side framework

2. MySQL:

MySQL is a well-known open-source DBMS (data set administration framework). MySQL is not difficult to utilize, solid, and quick. A DB the board framework that deals with implanted frameworks as well as client-server frameworks.

VI. EXPERIMENTAL SET UP AND RESULTS

A. Procedure

i. Web Scraping:

Step 1: To scrap the data firstly we are invoking the web driver of chrome by using Selenium library in python.

Step 2: Load the required website using that web driver.

Step 3: Extract the html code to string format from the website by using the Beautiful Soup.

Step 4: Extract the required data like tollgate name, location, cost etc by using some string operation.

Step 5: Finally, store the extracted data in a required format such as csv, json etc.
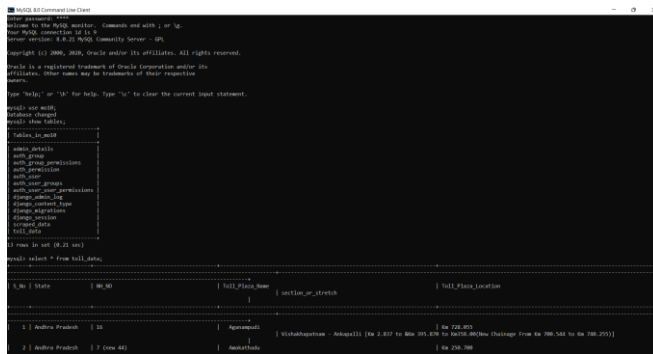

Fig 2: Scraped Data

Fig 2 Description: The above image shows the extracted data being stored in the database

ii. Admin:

Step 1: To obtain toll data in India we scrap data from NHAI website.

Step 2: After scraping the data we directly store the data in a Database (MySQL).

Step 3: Also, the data is scrape periodically so that all the updates made in NHAI website are applicable in our website.

iii. Client:

Step 1: To display the tolls and calculate the toll cost we are collecting way points from source to destination by using Google Maps API.

Step 2: Pointing Upper bounds and lower bounds with respective to the path called as polygon area.

Step 3: Checking each tollgate location whether it is present on the polygon area or not.

Step 4: If it is present on the polygon area, then we can call that tollgate is located on the path and displayed on Map.

Step 5: After displaying each toll, then itself we are storing the toll cost into a variable.

Step 6: By adding all toll costs at each toll to the variable, finally we will get total toll cost between source and destination.
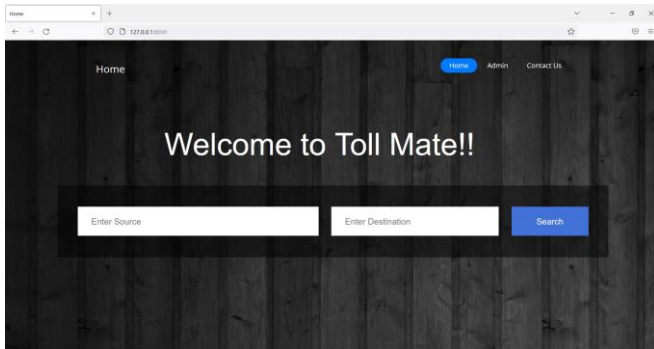
B. Results

Fig 3: Admin Home Page

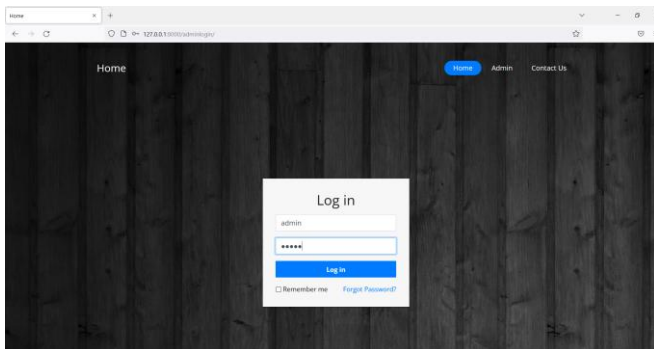Fig 3 Description: This is the starting page that is displayed to all the users.

Fig4: Admin Login Page

Fig4 Description: Admin must Signin in order to scrap the data.
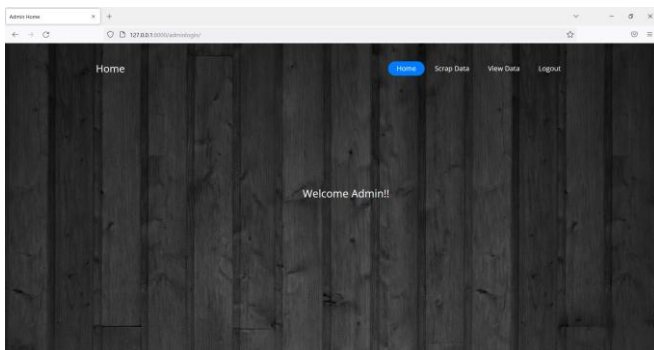

Fig 5: Admin Dashboard

Fig 5Description: After logging in, admin has the privilege to scrap the data and that scraped data will be stored into MySQL Database.
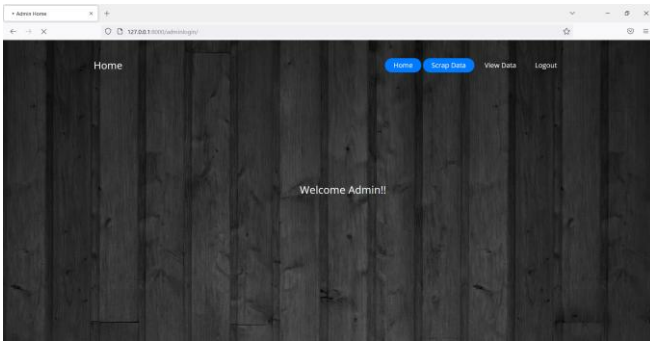
Fig 6: Scrap Data

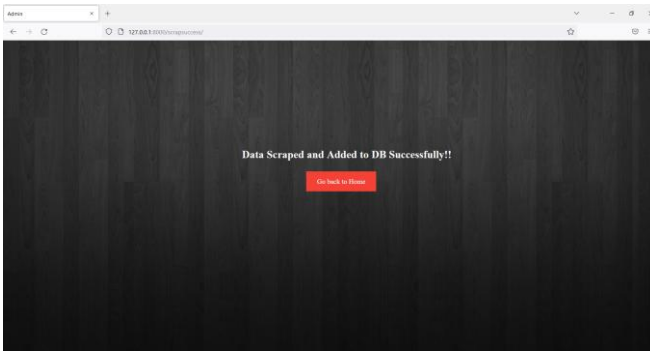Fig 6Description: When Admin clicks on Scrap Data the data will be scraped and will be stored into the Database.

Fig 7: Data Scraping Success

Fig 7Description: This page is displayed when scarping and storing into the database operations are successfully executed.
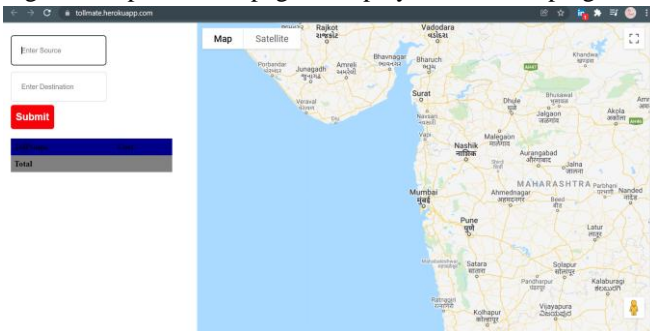

Fig 8: Estimate Toll Cost

Fig 8 Description: User must enter the source and destination of his/her travel and then click on submit.
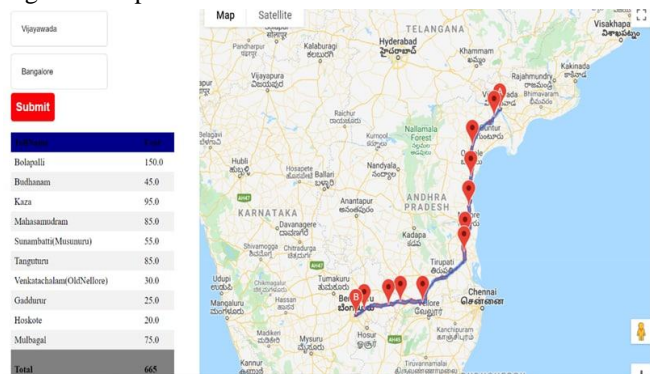

Fig 9: Display Toll Cost

Fig 9 Description: When Vijayawada and Bangalore are given as source and destination all the tolls between that path are displayed along with the total toll cost estimated.

VII.CONCLUSION

Finally, we have developed a web scraping algorithm for web scraping the govt NHAI toll data website and the data will be periodically updated automatically. Then we developed polygon bound algorithm for plotting the tolls on the map for the

optimized route and calculated the total toll cost for the journey without missing even one toll gate. Based on the toll cost,our app will recommend the users to recharge their FASTag wallet.

VIII.FUTURE SCOPE

In this thesis we mainly focused on Toll cost Estimation for only one country i.e., INDIA, but in the future, it can be expanded to all parts of the world. Also, while scraping data from NHAI (National Highway Authority of India) website it originally took half an hour to scrap and store it in thedatabase and after optimizing the scraper code it is now taking a 10-minute interval to completely store and scrap the data into the Database.So in the future, we can make sure that scraping and the storing of data can be done in seconds of time or maybe below a 5-minute interval.

REFERENCES

[1] Priya Matta, Nikita Sharma, Devyani Sharma, Bhasker Pant, Sachin Sharma, "Web Scraping: Applications and Scraping Tools", 2020.

[2] S. Amrin, "An Introduction to Fastag: A game changer in automatic toll collection systems in India", 2019.

[3] Esat Dere, Akif Durdu, "Usage of the A* Algorithm to Find the Shortest Path in Transportation Systems", 2018.

[4] Yifan Chen, "A flow model for optimization of highway toll stations", 2017

[5] MO Karpushko, I L Bartolomei, E N Karpushko, A V Zhidelev, V I Bryzgalov, "Analysis of Efficiency of Toll Road Network Development", 2021.

[6] Vyacheslav Kalashnikov, José G. Florez-Muniz, Nataliya Kalashnykova, Vladik Kreinovich, "A Heuristic Solution of the Toll Optimal Problem with Congestion Affected Costs",2016.

[7] Zhaoying Li, Zhao Zhang, Hao Liu, Liang Yang, "A new path planning method based on concave polygon convex decomposition and artificial bee colony algorithm", 2019.

[8] NevenaVajdic, Goran Mladenovic, Ceaser Queiroz, "Estimating minimum toll rates in public private partnerships", 2012.

[9] A. Renugambal, V. Adi Lakshmi Kameswari, "Finding Optimal Vehicular Route Based On GPS", 2018.

[10] Huiping Liu, Cheqing Jin, Bin Yang, Aoying Zhou, "Finding Top-k Optimal Sequenced Routes",2018.

[11] Moaiad Ahmad Khder, "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application", 2021.

[12] Hamid Mirzahossein, Shahriar Afandizadeh Zargari, "A Combined Model of Congestion Toll Pricing Based on System Optimization with Minimum Toll", 2018.

[13] R.K. Yadav,GirirajKishor,Himanshu,KishanKashyap, "Comparative Analysis of Route Planning Algorithms on Road Networks", 2020.

[14] B. Yang,F. Liu,Y. Zhang, and J. Chen, "The Modelling and the Optimization of Toll Plaza in Toll Station,2017.

[15] J. Fan,H.He,T.Hu,G.Li,L.Qin, and Y.Zhou, "Rasterization Computing-Based Parallel Vector Polygon Overlay Analysis Algorithms Using OpenMP and MPI", 2018.