# GTFS-Viz: Tool for Preprocessing and Visualizing GTFS Data

**Narumon Kunama**
Dept. Computer Engineering

Faculty of Engineering

Chiang Mai University, Thailand
narumon_kunama@cmu.ac.th


**Mudtana Worapan**
Dept. Computer Engineering

Faculty of Engineering

Chiang Mai University, Thailand
mudtana_w@cmu.ac.th


**Santi Phithakkitnukoon\***
Excellence Center in Infrastructure
Technology and Transportation
Engineering (ExCITE) and
Dept. of Computer Engineering

Faculty of Engineering

Chiang Mai University, Thailand
santi@eng.cmu..ac.th


**Merkebe Demissie**
Department of Civil Engineering

University of Calgary, Canada

merkebe.demissie@ucalgary.ca


*\*Corresponding author*

## Abstract

This paper presents a tool called *GTFS-Viz* for preprocessing and visualizing General Transit Feed Specification (GTFS) data. GTFS data is a common format for the public transportation schedules with associated geographic information. The tool consists of two main parts: preprocessor and visualizer. The preprocessor retrieves and transfers the GTFS data into a new data format for the visualizer. The visualizer displays an animation of public transport movement according to its schedule specified in the GTFS data on a map. It also displays a graph of the number of public transport vehicles per hour, and allows the user to choose a specific transport route, trip, and stations to visualize. This is the first attempt to build a tool that preprocesses and visualizes GTFS data that is publically available in a format that is not so easy to process with existing standard data visualization tools.

## Author Keywords

GTFS data; visualization; General Transit Feed Specification

## ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

## Introduction

The General Transit Feed Specification (GTFS) defines a common format for public transportation schedules and associated geographic information such as station locations. GTFS allows the transit data to be made available in a standard format that can be used for analysis by other engineers and researchers, or building useful applications by the developers. The GTFS feeds are a series of text files that are compressed into a ZIP file that include the information about fixed-route schedules, routes, and bus/train stations. GTFS datasets have been used to develop a variety of applications, such as trip planners, mobile applications, timetable generation software, tools for transit planning and operation analysis, and other categories of applications outlined in [1]. Visualizing the GTFS data can be useful for visual or qualitative analysis of the transit system of a particular area or city. Although, the GTFS data are in a common format, which are publically available for download, it is not quite straightforward to handle the data. In general, the data is handled subjectively by individuals. There is no tool available to visualize the GTFS data. Individuals who deal with the GTFS data must have a decent knowledge background in information technology in order to understand the GTFS data structure and be able to extract or mine the data accordingly. Transport engineers, urban planners, and researchers in other related fields can also get benefit from analyzing the GTFS data. To address this, we developed a tool called *GTFS-Viz* that processes and visualizes the GTFS data for exploratory analysis.

## Related work

Since the GTFS data provide useful information from which more insights can be gained when analyzed it with additional data sources e.g., shapefiles, traffic and delay information, and so on. Raghothama et al. [2] analyzed GTFS data from two cities; Rome and Stockholm, to gain insight into the sources and factors influencing public transport delays in the cities. Their analysis was carried out using MongoDB[1] and PostgreSQL[2] for data preprocessing, and QGIS[3] for data visualization (with static graphics). To rerun their analysis, one would need to have knowledge in information technologies, especially the ones mentioned for preprocessing and visualizing the GTFS data. Our developed tool, *GTFS-Viz*, can be used to simplify the procedure of preprocessing and visualizing the data – not only static type graphics but interactive and dynamic graphics. Bast et al. [3] developed an online real-time visualization of public transit based on GTFS and GTFS-realtime data. The GTFS-realtime provides information on delay, however it is not widely available to much larger number of developers. The work in [3] is useful for checking the transit connections and real-time delay information. However, the system does not allow user to upload GTFS data of any particular city to visualize. Our developed *GTFS-Viz* addresses this issue.

## GTFS Data

A typical GTFS feed contains a collection of text files that describes static public transit schedules and related geodata as following:

---

[1] https://www.mongodb.com

[2] https://www.postgresql.org

[3] http://www.qgis.org

- *agency.txt* –It contains information about the service agency of the feed, and time zone of the city where the transit operates.
- *stops.txt* –It contains the transit station/stop names, stop ID, and geo-locations (latitude, longitude).
- *trips.txt* –It contains the directions headed of each vehicle movement. Each trip has a service ID which specifies the days it operates on.
- *route.txt* –It contains the information on how trips are grouped into single services.
- *stop_times.txt* –It contains the information on the arrival and departure time of each transit station/stop, as well as station sequence and corresponding trip IDs.
- *calendar.txt* –It contains the information on weekly service days referenced by trips.
- *calendar_dates.txt* –It contains the information about weekly services that extend *calendar.txt* with some exceptions.
- *shapes.*txt –It contains geographical polylines representing the routes that a transit vehicle takes.

The relationships among the above text files are illustrated in Fig. 1. GTFS data from different providers have been made publically available on websites such as Transit Feeds[4] and GTFS Data Exchange[5].
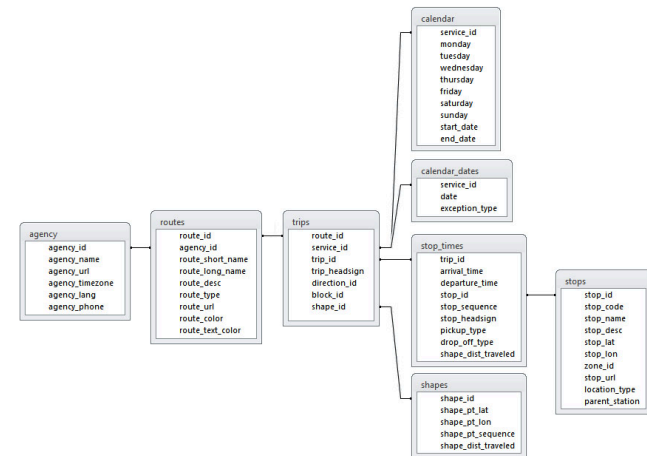


**Figure 1:** Relations among different text files of a GTFS feed.

## Data Preprocesser

To visualize the GTFS, those text files (Fig. 1) must be preprocessed and transformed into a file format that can be conveniently read by the visualizer engine. Since a typical data visualization tool is not designed to heavily deal with a database query or large-scale data preprocessing, but rather designed to support graphics rendering functionalities. Our visualizer is built with the Processing Development Environment (PDE)[6], so we developed our preprocessor that takes the GTFS feed's text files and generates a CSV file containing necessary information required for the visualizer.

We built the GTFS data preprocesser using Java language that creates and connects to a database in

4 http://transitfeeds.com

5 http://www.gtfs-data-exchange.com

6 https://processing.org

which a GTFS feed files are stored and processed. We created a user interface for the ease of use to retrieve the GTFS data from the database and then be transformed into a CSV file for the visualizer.
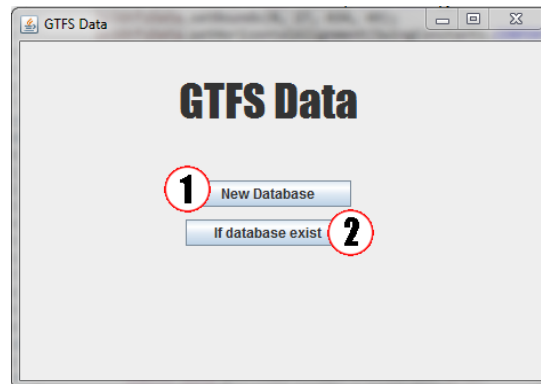
*Home screen*
Preprocessor's main screen is shown Fig. 2 where each button is described as following:
  (1) New Database: To create a new database. This will take the user to Fig. 3.
  (2) If database exists: If the database exists, the user can skip the process what creates a new database, which will take the user to the page shown in Fig. 5.
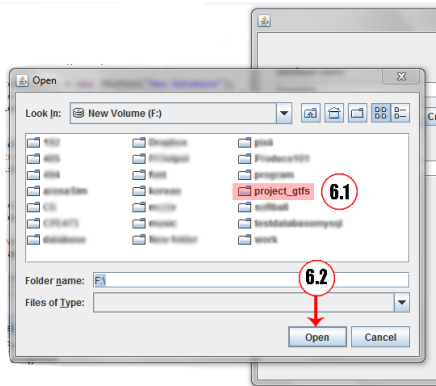


**Figure 4:** A snapshot showing a path where to gather the GTFS data files into database (6.1). Click "Open" (6.2) to collect the file into database.



**Figure 2:** A snapshot of the home screen.

*New database*
As shown in Fig. 3, the user gives a name to the new database if the database does not already exist. Buttons are described as following:
  (1) Database name: Type in database name.
  (2) Create: To create a database with the specified name.

(3) Open: To choose the path to gather the GTFS data into the database (Fig. 4).
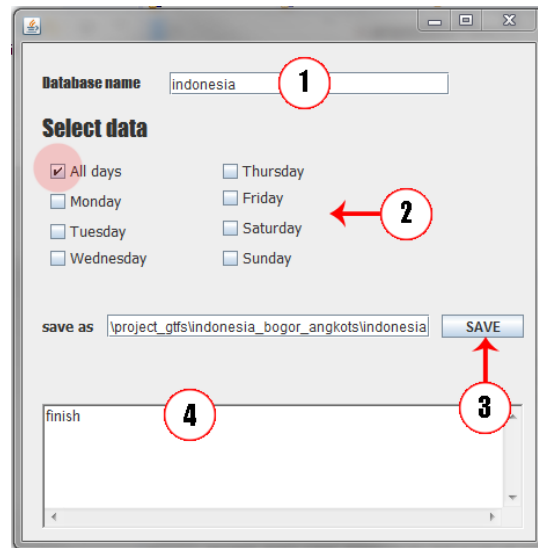(4) Next: Go to the page that allows the user to retrieve the data from the database.
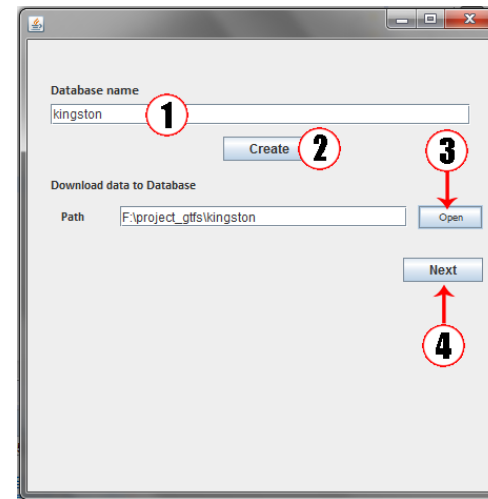


**Figure 3:** A snapshot of the page for creating a new database.

*Retrieving data from database*
At this stage, the preprocessor will be transforming the GTFS feed files into a CSV ready to be retrieved for visualization. Buttons in Fig. 5 are described as following:
  (1) Database name: Type in a database name that the user wants to retrieve the CSV file for visualization.
  (2) Select data: The user can select weekly data to retrieve.
  (3) Save: The user chooses where to save the output file.

(4) Status box: To show the status of data processing.



**Figure 5:** A snapshot of the interface that allows the user to retrieve the output CSV file from selected the database.
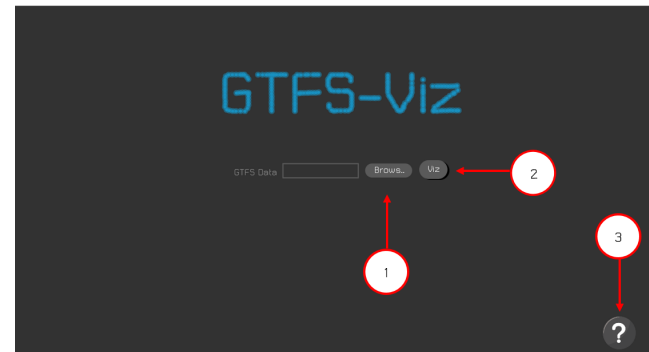
## Data Visualizer

Once the CSV file has been generated and retrieved from the preprocessor, the *GTFS-Viz*'s visualizer engine can fetch it and render the graphics in a graphic interface developed with the PDE.

*Home screen*

The home screen is shown in Fig. 6 where the user can select which GTFS data to visualize.

(1) Browse: To bring the user to the folder where the preprocessor's output file (CSV) of the GTFS feed is located.
(2) Viz: To visualize the selected file on an interactive map.
(3) Question mark icon: To show the user manual, as shown in Fig. 7.
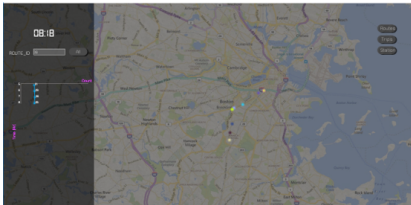(4) Back (Fig. 7): To go back to the main screen.



**Figure 6:** A snapshot of the visualizer's home screen.
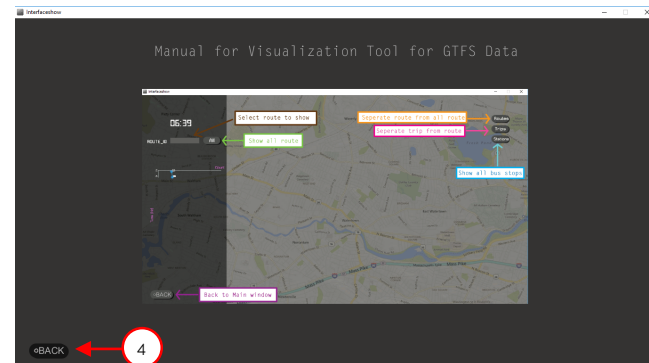


**Figure 7:** A snapshot of the user manual.
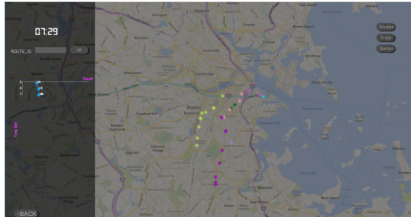
**Figure 9:** Vehicles are colored according to their routes.


**Figure 10:** Vehicles are colored according to their trips.

*Visualization*

Once the user clicks on the "Viz" button in Fig. 6, it brings the user the visualization interface (shown in Fig. 8) that displays the GTFS feed on an interactive map along with a display control panel.

(1) A reference running time for the visualization.
(2) The user can choose individual or all routes to visualize.
(3) Graph showing the hourly amount of vehicles.
(4) To bring the user back to the home screen. This is normally provided for the user to go back and select a different GTFS feed to visualize.
(5) This is where the action is. Movement of transit vehicles are shown on an interactive map that the user can zoon in and out.
(6) The user can choose to view the differentiation (shown with different colors) of routes (Fig. 9) and trips (Fig. 10), and the station locations.
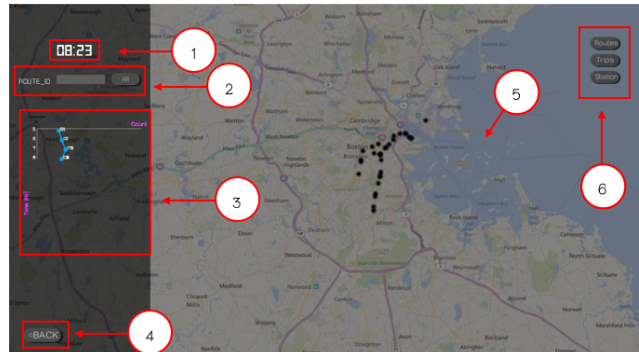


**Figure 8:** A snapshot of the visualization interface.

**Demo**

For the demonstration purposes, a video clip showing how the *GTFS-Viz* works is available at: https://youtu.be/snkfiXx8w3U.

**User Experience Study**

To evaluate the *GTFS-Viz* in terms of the user experience, we brought it to the real users to try out. We conducted a user experience study with a total of 50 participants that included 25 males and 25 females. Each participant was asked to use the tool and then asked to answer a questionnaire. The questionnaire was designed based on the Theory of Four Elements of User Experience [4]. Participants were asked to give a rating score according to their level of agreement with the following four statements regarding their experience with the *GTFS-Viz*:

1. It is useful.
2. It is easy to use.
3. It is easy to start using.
4. It is fun and engaging.

The rating is a 5-likert scale where 1 means the lowest level and 5 implies the highest level of agreement. Along with the four statements, the participants can also give comments and suggestions at the end of the questionnaire form, as shown in Fig. 11.

The overall result is shown in Fig. 12. The aspect of the system is being easy to start using is on average rated highest (4.58) while the easy-to-use is rated lowest (3.92). One of the comments from a business owner (in 30-39 age group) was "*There should be a clear system manual provided. It will make the system more easy to use*". This result suggests that we need to improve in the aspect of easy-to-use in our future development.

**Figure 11:** Survey questionnaire.
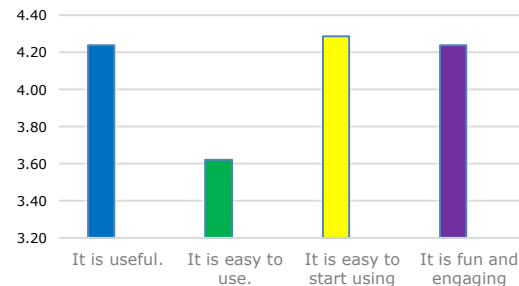


**Figure 12:** Overall user experience study result.

When grouping the participants' response by gender (Fig. 13), we found that male users gave the highest

rating to the system being useful (4.30) and gave the lowest rating to the easy-to-use aspect (3.80) of the system. On the other hand, female users highly rated the easy-to-start-using aspect (4.55) while gave the lowest rating to the easy-to-use (3.45) similar to the overall result. One of the male participant commented "*The system is very useful, but the graphics need improvement*". One of the female users commented "*It's not very easy to use but I think it may not be so difficult to learn how to use it. I can see myself using it. It is useful that it helps me see how the transit looks like in different cities around the world*". The female users on average think that it is easy for them to start using the system than males, while males think that the system is easier to use than the females.
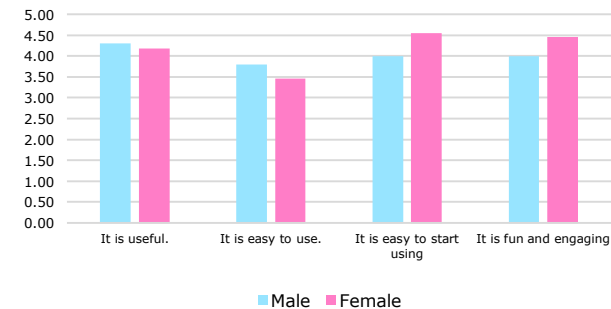


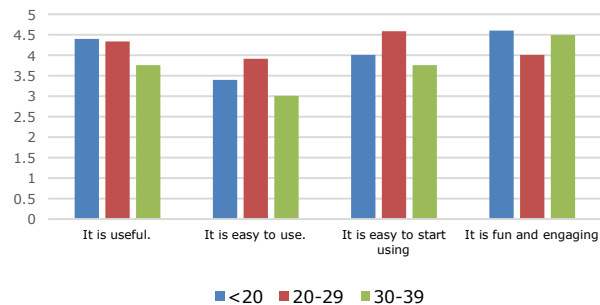**Figure 13:** User experience study result when grouped by gender.

We further grouped the participants result by age group (Fig. 14). Five participants were less than 20 years old. There were 40 participants in the 20-29 age group, and five participants were in the 30-39 age group. The result (Fig. 14) shows that the users who are less than 20 years of age think the system is fun and engaging

the most (4.60), but gave the lowest rating to the aspect of the system being easy to use (3.40). The users who are in 20-29 age group highly agree that it is easy for them to start using the system (4.58) while least think that the system easy to use (3.92). The users of age group 30-39 gave the highest average rating to the aspect that the system is fun and engaging (4.50), while the easy-to-use aspect was given the lowest average rating (3.0).

A student who was less than 20 years old commented, "*The graphic looks interesting. It would be nice if we (the user) can adjust the speed of the animation*". Another student from the 20-29 age group commented, "*It is a useful tool. I like it. But the graph should look clear. It's difficult to read the graph*". A company worker who was in the 30-39 age group commented, "*It is a fun tool. I can see transit in other cities around the world*". Another user who was a researcher in transport engineering in the 30-39 age group commented, "*The graph should be more clear. Instructions should be provided. I think it's useful and there are still rooms for improvement*".



**Figure 14:** User experience study result when grouped by age.

The participants provided very valueable comments from which gave us ideas for our future development, such as user manuals, clearer graph, and speed control.

## Conclusion

GTFS is a general data format that describes static public transit schedules and related geodata. A typical GTFS feed data consists of a set of text files each describes different characteristic of the public transit such as trips, routes, and sequences. Being able to visualize the GTFS data is essential for exploratory analysis of the transit system and other related domains e.g., urban planning, traffic management, and human geography studies. General visualization tools are not designed to cope with database queries and data preprocessing, thus it becomes a barrier for whom interested in visualizing the GTFS data. Inspired by this issue, we developed a tool for preprocessing and visualizing the GTFS feed data with an easy to use interfaces, called *GTFS-VIz*. The developed system is composed of two main components; preprocessor and visualizer. The preprocessor was built with Java that handles database connection and file transformation for visualization. The visualizer was built with Processing Development Environment (PDE) that renders the GTFS data on an interactive graphics and animation. View options and statistical graph are available on the interface. Our preliminary user experience study suggests that there needs to be an improvement on the easy-to-use aspect of the system. Nonetheless, the system was well received overall. We believe that the *GFTS-Viz* is a useful application for people in transport and urban planning in general, as well as other related fields.

**References**

1. A. Antrim and S. J. Barbeau. The Many Uses of GTFS Data – Opening the Door to Transit and Multimodal Applications. In *Proc. ITS America's 23rd Annual Meeting & Exposition*, 2013.

2. J. Raghothama, V. Magal Shreenath and S. Meijer. Analytics on Public Transport delays with Spatial Big Data. In *Proc. ACM BigSpatial*, 2016.

3. H. Bast, P. Brosi and S. Storandt. Real-Time Movement Visualization of Public Transit Data. *In Proc ACM SIGSPATIAL*, 2014.

4. F. Guo. 2012. More Than Usability: The Four Elements of User Experience. Retrieved April 22, 2017 from http://www.uxmatters.com/mt/archives/2012/04/more-than-usability-the-four-elements-ofuserexperience-part-i.php