BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023

http://excel.fit.vutbr.cz

# Analysis and Visualisation of Brno Public Transport Data

Juraj Lazúr*

**Abstract**

The planning and management of modern public transport systems is not possible without real operating data. The aim of this work is to simplify the procedure of extracting useful knowledge from data on vehicle behaviour in the Brno public transport system. The solution lies in the design of a tool that stores data from real operation and allows its analysis over time. The basis is a regularly updated system model, over which data from real operation is processed into a format designed for subsequent analyses. The result is a system that autonomously monitors changes and behaviour of the transport system over time, and provides a platform for analysis. The proposed solution can mirror the transit system with 92% accuracy and reduces the volume of stored data by 75% on average. The implemented tool can serve, for example, to detect bottlenecks or the probability with which a given connection will be delayed. The processed data can also serve as input for further advanced analyses. At the same time, the proposed solution is reusable for other transit systems described by the GTFS format.

*xlazur00@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

## 1. Introduction

Public transport systems keep everyday life moving. Properly designed and working systems can save **costs and time**. However, the effectiveness of the design and running of these systems depends on knowledge of their **actual behaviour**.

Extracting knowledge from data describing real system behaviour is a non-trivial task. A potential solution must be able to respond to **frequent changes**, allow monitoring of their impact on the system and detect bottlenecks. At the same time, it must be able to deal with the anomalies that real data contains. The added value then lies in making this knowledge available to the **general public**.

The majority of currently available tools work primarily with the current state of the system. Tools such as TRAVIC[1][1] or Interactive Map of IDS JMK[2] clearly visualise the **current state** of transport systems, but do not track the behaviour of the systems over time. The second category tools such as Babitron[3] use statistical data, but do not provide space for deeper analysis. The third category tools such as Explore Brno's Public Transport System[4] can extract knowledge only from **static data**.

My proposed tool, which was created in cooperation with **the Municipality of Brno**, records the behaviour of the Brno public transport system over time and provides a platform for analysing this behaviour. The solution includes methods for modeling the system from **GTFS** format and processing data from real operation.

The implemented solution provides a simple possibility of **gaining knowledge** from the behaviour of the Brno public transport system. The proposed system model is able to react flexibly to changes and can serve as a source for **deeper analyses**. The problematic aspects of the solution are the time required for data processing as well as the imperfections in the mirroring of the real system.

[1] https://travic.app
[2] https://mapa.idsjmk.cz
[3] https://kam.mff.cuni.cz/~babilon/zpmapa
[4] https://mestobrno.maps.arcgis.com/apps/MapJournal/index.html?appid=6ce3a916606e4c1996015613ef57e016

## 2. Design of the Solution

The proposed solution consists of **three units**. The **first** unit, described in section 2.1, is the concept of accessing vehicle behaviour records. This section also describes and evaluates a method for record processing. The **second** unit described in section 2.2 contains the overall architecture of the proposed tool, together with the internal model of the traffic system. The **third** unit containing a brief description of the analysis platform and the implemented analysis methods is described in section 2.3.

### 2.1 Input records

**Vehicle records** are the basic data source of the whole tool. Each vehicle with a logged-in driver sends a record of its current status at 10-second intervals, which includes location, delay and other data. However knowledge extraction from raw data is inefficient.

The proposed solution transforms the data into a format that describes the progress of **individual trips** in a specific time and space. Thus, the processing of input records is based on grouping and assigning records based on their geographical component to specific paths of the trip route. This method is illustrated graphically by the Figure 2.1.

The transformation with respect to the trip was chosen because of the structure of the transport system itself, where the trip represents the lowest level of **abstraction**. The advantages of such a solution are the reduction of redundant data, the ability to aggregate data, the ease of modeling the relationships between trips, and the ability to easily extract knowledge from these records. The disadvantages are the time and space complexity of the transformation and the risk of losing interesting anomalies.

### 2.2 Tool architecture

The core of the tool is a **model of the transit system**, which serves as the basis for processing records of vehicle behaviour. This model contains stops, lines, routes and trips data. At the same time, the model responds to the changes that the transport system undergoes over time. Thus, the processing of the records always uses the state of the transport system that is valid at the time the records are created.

The data source for the calculation of the model is **GTFS** data, which is not complete. These data do not contain definitions of routes in space. The solution is a custom routing algorithm that computes the route definitions. The overall scheme of the tool, along with the data sources and implemented methods, is visualized by the Figure 2.2.

### 2.3 Analytical platform

Its main task is to generate outputs that simplify **the extraction of knowledge** from processed data about system behaviour. The basic function is the aggregation and visualization of processed data from a selected time period.

Two basic analytical functions are implemented within the system. The first one analyses the probability of delay of a specific trip, while the second one analyses the evolution of the delay during the trip.

## 3. Results

The result of this work is a system capable of monitoring and recording the behaviour of the Brno public transport system.

The implemented tool mirrors the transit system with an accuracy of **around 92%**, with inaccuracies due to both imperfect routing and imperfect input GTFS data. The accuracy of the routing itself depends on the type of transport, as shown in Section 3.1.

The proposed record transformation reduces the stored data by **75% on average**, with the reduction rate again mainly depending on the transport type, as shown by the plots in Section 3.2.

The processed data is used by the implemented analytics platform, whose interface is shown in Section 3.3. It is this platform that generates the output that is intended to simplify the extraction of knowledge from the data about the behaviour of the Brno public transport system.

Further development should be directed mainly towards the implementation of **further analysis methods** and testing the tool on other transport systems defined by GTFS format data.

## Acknowledgements

I would like to thank my supervisor Ing. Jiří Hynek, Ph.D. for his invaluable advice, time and motivating words when my work was stuck. I would also like to thank Mgr. Martin Dvořák and Mgr. Jan Zvara, Ph.D. from the Municipality of Brno for their helpful advice and critical insights.

## References

[1] Hannah Bast, Patrick Brosi, and Sabine Storandt. Real-time movement visualization of public transit data. In *Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 331–340, 2014.