Public Transit Network File Format

Tobias Zündorf
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
zuendorf@kit.edu

August 8, 2017

1 File Overview

Each public transit network consists of five files (`footpaths.gr`, `footpath.co`, `stops.csv`, `connections.csv`, `trips.csv`). The first two files describe the footpath graph of the network using a DIMACS graph format[1]. The remaining three files contain the timetable of the public transit network using a simple csv format. Note that the format of all files is text based. Also note that the indices of all entities start with 1 and not with 0.

2 Footpath Graph

The footpath graph is a directed and weighted graph $G = (V, E)$ and is defined using two files: The first file contains the edges of the graph (`footpaths.gr`). The second file contains the vertices with their coordinates (`footpath.co`). Both files use the DIMACS format.

2.1 footpaths.gr

This file contains all edges of the footpath graph. The first line of the file contains information about the size of the graph and has the following format:

`p sp` $\langle n \rangle$ $\langle m \rangle$

Where $n = |V|$ is the number of vertices and $m = |E|$ the number of edges. The remaining lines of the file define the edges of the graph (one edge per line). The format of these lines is

`a` $\langle u \rangle$ $\langle w \rangle$ $\langle \tau \rangle$,

which represents an edge from the vertex $u \in V$ to the vertex $v \in V$. The travel time required to walk from u to v is τ and is specified in seconds.

2.2 footpath.co

This file contains the geographic coordinates of the vertices. The first line of the file has the format

`p aux sp co` $\langle n \rangle$,

where $n = |V|$ is again the number of vertices in the graph and therefore matches the number of vertices as specified by the `footpaths.gr` file. Each of the remaining lines defines the coordinates of a vertex using the format

`v` $\langle v \rangle$ $\langle lon \rangle$ $\langle lat \rangle$,

[1] http://www.diag.uniroma1.it/challenge9/format.shtml#graph

where $v \in V$ is the vertex, *lon* the longitude, and *lat* the latitude.

3 Timetable

The timetable is specified using a file format that reflects the data structures required for the connection scan algorithm (CSA) [1]. The definition of the timetable consists of three files: The first file describes the stops $S \subseteq V$ of the network (`stops.csv`). The second file describes the connections (`connections.csv`). The third file contains additional information about the trips (`trips.csv`). All three files are csv-files and use the comma character (`,`) to seperate columns.

3.1 stops.csv

This file defines the stops of the public transit network and has three columns

`vertex_id`,
`gtfs_stop_name`,
`minimum_change_time`.

The stops of the public transit network are a subset of the vertices of the footpath graph ($S \subseteq V$). Accordingly, the first column (`vertex_id`) contains the vertex that corresponds to the stop. The second column (`gtfs_stop_name`) contains the name of the stop as it was extracted from the GTFS[2][3] feed. The third column (`minimum_change_time`) specifies the minimum time required to change between different trips at this stop (in seconds).

3.2 connections.csv

This file defines the connections of the network and has six columns

`connection_id`,
`departure_vertex`,
`arrival_vertex`,
`departure_time`,
`arrival_time`,
`trip_id`.

[2]https://developers.google.com/transit/gtfs/reference/
[3]http://gtfs.geops.ch/

Each row of the file (except for the header row) defines one connection of the timetable. The departure and arrival stop of the connection are specified in the `departure_vertex` and `arrival_vertex` columns. As all stops are vertices, the stops are defined using the corresponding vertex id. The departure and arrival time of the connection are specified in the `departure_time` and `departure_time` columns. Both values are absolute numbers, representing seconds since 00:00 am of the first day of the timetable. Finally, the trip that the connection belongs to is specified in the column `trip_id`. If two connections are part of the same trip (share a common trip id) then they can be used in one journey without requiring a transfer between vehicles. Furthermore the trip id can be used to obtain additional information about the trip and the vehicle.

3.3 trips.csv

This file contains additional informations about the trips in the network. However, this data is not required for the journey or profile computation. The file has five columns

`trip_id`,
`gtfs_trip_short_name`,
`gtfs_route_type`,
`gtfs_route_short_name`,
`gtfs_route_long_name`.

Each row provides supplementary information about the trip specified by the `trip_id`. Amongst others this includes the vehicle type and the name of the route that the trip belongs to. All columns correspond to the data of the original GTFS feed that was used to obtain the network.

References

[1] Julian Dibbelt, Thomas Pajor, Ben Strasser, and Dorothea Wagner. Intriguingly Simple and Fast Transit Routing. In *Proceedings of the 12th International Symposium on Experimental Algorithms (SEA'13)*, volume 7933 of *Lecture Notes in Computer Science*, pages 43–54. Springer, 2013.