

Programmeringsprojekt SL-light

Beskrivning av dataset

Patrick Wentzel, DSV ACT, februari 2020

För den som vill testa/utveckla sin lösning med verkligt data och se hur lösningen fungerar med lite fler noder finns nu ett förberett dataset.

De facto-standarden för den här typen av transportdata (för kollektivtrafik o dyl) heter General Transit Feed Specification (GTFS) som består av ett antal csv-filer som kan ses som tabeller i en normaliserad relationsdatabas som kopplas samman via primär- och främmande nycklar.

För en översikt om GTFS:

https://en.wikipedia.org/wiki/General_Transit_Feed_Specification

Referens för filerna:

<https://developers.google.com/transit/gtfs/reference>

Datasetet är en kraftigt reducerad delmängd av ett dataset som omfattar hela Sveriges kollektivtrafik och är minskat så att det bara innehåller data för SL, specifikt tunnelbana, spårväg och bussar med nummer under 100 (dvs innerstaden) för en enda dag (2020-01-31).

En enkel beskrivning av hur datat hänger ihop

Transportörer (*agency*)

→ har rutter (*routes*)

(→ som trafikeras enligt ett schema (*calendar*))

→ som består av turer (*trips*)

→ som består av en sekvens av stopptider (*stop_times*)

→ som sker vid hållplatser (*stops*)

Att bygga en graf med GTFS-data

Det finns många sätt att bygga upp en graf med det här datasetet, men de viktigaste delarna finns i filerna **stops.txt** som innehåller alla hållplatser (med koordinater) och **stop_times.txt** som innehåller information om avgångar. Även **routes.txt** och **trips.txt** kan innehålla intressant data.

För det aktuella datasetet saknar filerna **calendar.txt** och **calendar_dates.txt** helt betydelse eftersom datat bara gäller en enda dag.

Bildfilen graph.gv.png visar hur en graf byggd från datasetet kan se ut (genererad med dot).

Rader i filerna:

1	sl_agency.txt
38	sl_routes.txt
143044	sl_stop_times.txt
443	sl_stops.txt
6679	sl_trips.txt
152185	total

Det är absolut inte ett krav att använda filerna utan de finns enbart för att man ska ha möjlighet om man vill.

Klassdiagrammet för modellen som GTFS använder (de viktigaste klasserna till vänster):

