Public transport applications based on data

from searching data sources to implementation

Marcin Domański 2016

KODUJ DLA POLSKI

HACKERSPACE SILESIA

Agenda

- Types of data in public trasport
- Possible data sources
- Case study: #jakniedojade (application based on data from private partner)
- Case study: popugraf (analysis base on open data from Eurostat)
- Case study: EasyRider (application based on crowdsourcing)
- Human resources tips for non-profit technological projects

Why public transport?

- We could reuse code since it is unlimited
- We should reuse space since it is limited
- Individual transport sector get more profits, so public transport needs more support from non-profit side
- By public transport I mean buses, tram, trains, but also going on foot and by bike ☺

Data needed for public transport applications

- Timetables
- Maps
- Population data
- Demand data (from measurements or crowdsourcing)
- Other geographical located data (e.g. real estate prices, salary data, employee data)

How to get high quality data?

- Standardization
  - Building universal application for each location
  - Using already developed functions
- Availability
  - Open public data…
  - Fight for opening public data
  - Cooperation with private companies, using their own databases
  - Hacking ;)

Timetables – potential sources

- Standards
  - GTFS (General Transit Feed Specification) – created by Google
    - Used in Google Maps route search
    - Not always opened by default, depends on public transport organizer decision. List is available here.
    - TRAVIC (real time visualization of GTFS data)
- Availability
  - jakdojade.pl („how I can get there”; Poland biggest connection searching engine for cities; private)
  - e-podróżnik.pl („e-traveler”; same as jakdojade.pl, but for long distance connections)
  - IDOS (Czech equivalent for polish e-podróżnik; supported by government; including all operators)
  - API for data!

Maps

- For analysis (used as data)
  - If we want to calculate something (total lenght of network, average distance, central points etc.)
  - OSM (OpenStreetMap; vector maps in xml format; challenges)
  - OCM (public transport on OSM)
  - Public transport organizers' own systems (standardization)
- For visualization
  - Customizing well known maps:
    - Google Maps
    - OpenStreetMap (uMap)
  - Output from analysis
  - Objects created directly on maps, like for example in Brand New Subway or Get Remix

Case study: #jakniedojade

- Under development ;)
- Cooperation with jakdojade.pl (project based on their searching results)
- Jak NIE dojadę (how I can't get there) vs. jak dojadę (how I can get there)
- From searching results we extract most popular routes with more than one transfer

#jakniedojade – data analysis

- Input as CSV file
- Conversion to format readable by GIS service (we use Quantum GIS)
- Clustering – agregetion stops into bigger groups
- Counting all connections between groups
- Sorting intercluster connections by popularity
- Some manual analysis (checking what the cluster is, for example: University campus or big settlement)

#jakniedojade – happening/website

- Most popular unconvienet routes (more than one transfer) are described and chosen for publication on website
- Voting for most uncomfortable route will be possible from chosen data
- Input for public transport organizer. It is much easier to make some changes in their network or/and timetables based on filtered data.
- Remember that this kind of analysis finds only problems of jakdojade.pl users, but this is very popular platform all across the Poland (expecting older people)

Casestudy: popugraf

- It’s my own name ;) Population + graph ;)
- Motivation: suburan rail on existing railways – where to place new stops?
- Special chart prepared based on population density data
- How many persons live in some distance from communication line
- Mapping population to distance
- Population data from Eurostat (rectangle grid 1km x 1km for whole UE)

Easyrider

- Crowdsourcing application basen on CyclePhilly idea.
- Measuring cycle paths popularity. Where to build new ones?
- Data gathering in the background: user should only install application (opposite as in all application like Strava or Endomodo)
- Automatic discovering of mobility type (pedestrian, bicycle, car)
- Strava heat map – without legend and values is only a curiosity

Questions?