

AUTOMATICKÉ ZPRACOVÁNÍ ČESKÝCH JÍZDNÍCH ŘÁDŮ AUTOBUSŮJan MASOPUST¹

¹ Katedra geoinformatiky, Přírodovědecká fakulta, Univerzita Palackého v Olomouci,
17. listopadu 50, 771 46, Olomouc, Česká republika
jan.masopust01@upol.cz

Abstrakt

Data z jízdních řádů veřejné hromadné dopravy jsou základem pro analýzy ve veřejné dopravě. V České republice jsou k dispozici data ve formátu JDF. Tato data nelze jednoduše přímo použít pro analýzy a data mining. Proto byla vytvořena aplikace pro automatické zpracování jízdních řádů veřejné dopravy ve formátu JDF. Díky technologiím Python, PostgreSQL a PostGIS aplikace převádí data do prostorové databáze a formátu GTFS, což je světový standard pro data jízdních řádů. Aplikace je open source a poskytuje data pro analýzu a optimalizaci veřejné dopravy. Použití těchto dat je možné v oblastech identifikace míst a časů zpoždění, optimalizace kapacity vozidel a návrhu sítě autobusových linek.

Abstract

Public transport timetable data are the base of analytics in public transport. In the Czech Republic there are data available in the JDF format. This data cannot be directly used for analysis and data-mining. We have created an application for automatic processing of public transport timetables in JDF format. Thanks to Python, PostgreSQL and PostGIS technologies, our application can convert data to spatial database and GTFS format, which is the world standard for timetables data. The application is an open source software and provide data for analysis and optimization of public transport – use of these data is in fields of identify places and times of delays, optimizing the capacity of vehicles and design network of routes.

Klíčová slova: jízdní řád; autobus; Python; GTFS; JDF

Keywords: timetable; bus; Python; JDF; GTFS

ÚVOD

V České republice zajišťuje svobodný přístup k jízdním řádům veřejné dopravy vyhláška o jízdních řádech veřejné linkové dopravy (122/2014 Sb.). Jízdní řády autobusové dopravy jsou volně dostupné ke stažení z celostátního informačního systému jízdních řádů (CIS JŘ) prostřednictvím FTP serveru *ftp.cisjr.cz*. Metodický pokyn Ministerstva dopravy (11/2014-190-CIS/6) určuje za formát pro autobusové jízdní řády jednotný datový formát JDF. Avšak světovým de-facto standardem v oblasti veřejné dopravy se stal formát General Transit Feed Specification GTFS (MobilityData, 2020). Kromě jeho široké rozšířenosti a množství aplikací pracujících s tímto formátem, patří mezi jeho hlavní výhody i podpora prostorových informací. Formát JDF s prostorovou složku nepracuje. Při podrobnějším zkoumání navíc vychází najevo, že jízdní řády v CIS JŘ obsahují mnohem méně informací než JDF produkované jednotlivými dopravci, koordinátory nebo krajskými úřady. Proto byl vyvinut program *jdf2gtfs*, který umožňuje konverzi jízdních řádů a zpracování historických dat a je volně dostupný na platformě GitHub (<https://github.com/masopust/jdf2gtfs>).

FORMÁTY JÍZDNÍCH ŘÁDŮ

Rozvoj technologií a nástup open source a big data se nevyhnul ani oblasti jízdních řádů. V místech, kde funguje volný trh v autobusové dopravě (např. Polsko), často nejsou k dispozici žádná strojově zpracovatelná data. Jednotliví dopravci vydávají jízdní řády ve formě HTML tabulek, textu, obrázků apod. V České a Slovenské republice, kde je trh autobusové dopravy zásadně regulovaný a řízený se zveřejňují a schvalují jízdní řády ve formátu JDF. Dlouhou dobu v Česku však nebyly tyto soubory volně dostupné, na Slovensku stále nejsou volně dostupné. V USA vznikl formát GTFS, který se rozšířil především do západní Evropy. Je na něm postavena řada prostorových i neprostorových aplikací (Zajíčková, 2018).

Jednotný datový formát JDF

Jízdní řád JDF je tvořen sadou textových souborů, které jsou zpravidla zabaleny do archivu ZIP. Textové soubory využívají formátu CSV se středníkem na konci řádku. Každý jízdní řád se skládá z devíti povinných souborů a až osmi nepovinných. Povinné soubory popisují zastávky, linky, spoje, časy odjezdů a příjezdů, dopravce, datumová omezení a další. Nepovinné soubory mohou přinášet informace o místenkách, označnicích, návaznostech apod. (Tabulka 1). Všechny soubory musí být kódovány v CP1250. V současné době je možné používat JDF ve verzích 1.9, 1.10, 1.11 (Metodický pokyn, 2014). V JDF není možné uvést žádnou prostorovou informaci ani o poloze zastávek ani o průběhu spojů a linek.

Vybrané odlišnosti mezi originálními soubory JDF a soubory JDF v CIS JŘ:

- V ČR existuje databáze zastávek s unikátním ID (*CIS číslo*), která však nepatří mezi data zveřejňovaná na CIS JŘ. Zde slouží jako unikátní identifikátor zastávky její název.
- Jednotlivé JDF soubory obsahují běžně všechny linky určité oblasti nebo dopravce. V CIS JŘ jsou tyto soubory rozděleny do jednotlivých JDF s pouze jednou linkou.
- Informace o označnicích jsou v CIS JŘ odstraněny.

Formát General Transit Feed Specification GTFS

Jízdní řád GTFS je stejně jako v předchozím případě tvořen sadou textových (CSV) souborů zabalených do archivu ZIP. Jde o formát pro jízdní řády a související geografické informace. Kromě statických informací podporuje také předávání realtime dat z provozu veřejné dopravy. Soubory musí být kódovány v UTF-8. Jízdní řád tvoří šest povinných souborů popisujících zastávky, linky, spoje, časy odjezdů a příjezdů, dopravce, datumová omezení. Další deset nepovinných souborů může dodávat např. informace o frekvenčních jízdních řádech, návaznostech a prostorovou informaci o průběhu linií – spojů a linek.

Databáze PostgreSQL a prostorová extenze PostGIS

Pro zpracování jízdních řádů, generování GTFS souborů a ukládání historických dat je vhodné použít databázi PostgreSQL s prostorovou extenzí PostGIS.

Tab 1. Srovnání a popis formátů JDF a GTFS (P = povinný soubor, N = nepovinný soubor)

GTFS		JDF	Popis
agency.txt	P	Dopravci.txt	P Seznam dopravců
stops.txt	P	Zastavky.txt	P Zastávky, GTFS obsahuje i polohu zastávek
		Oznacniky.txt	N V GTFS jsou zastávky i označnický v souboru stops.txt
routes.txt	P	Linky.txt	P Seznam linek
		LinExt.txt	N Příslušnost linky k MHD nebo IDS
		AltLinky.txt	N Alternativní čísla linek v jiných státech
-	-	Zaslinky.txt	P Seznam zastávek na jednotlivých linkách
trips.txt	P	Spoje.txt	P Seznam spojů
-	-	SpojSkup.txt	N Skupiny spojů - prezentace pod společným nadpisem.
stop_times.txt	P	Zasspoje.txt	P Časy odjezdů a příjezdů spojů na zastávkách
calendar.txt	(P)	Pevnykod.txt	P Provozní dny v týdny, Pevnykod - i jiné poznámky
calendar_dates.txt	(P)	Caskody.txt	P Výjimky z provozních dnů
fare_attributes.txt	N	-	- Informace o jízdném
fare_rules.txt	N	-	- Informace o jízdném
-	-	Mistenky.txt	N Informace o místenkách
shapes.txt	N	-	- Prostorové informace – linie tras spojů
frequencies.txt	N	-	- Frekvenční jízdní řád
transfers.txt	N	Navaznosti.txt	N Informace o přestupech a návaznostech
pathways.txt	N	-	- Informace o pěších přesunech mezi zastávkami
levels.txt	N	-	- Informace o podlažích ve stanicích
translations.txt	N	-	- Překlady názvů do cizích jazyků
feed_info.txt	N	VerzeJDF.txt	P Metadata
attributions.txt	N	-	-

AUTOMATICKÉ ZPRACOVÁNÍ JÍZDNÍCH ŘÁDŮ

V současné době existuje několik projektů, které se věnují konverzi jízdních řádů. Mezi nejpropracovanější patří *Aplikace strojového čtení souborů JDF* (Papež, 2020), která ale slouží ke konverzi do formátu XLSX a archivaci historických jízdních řádů. Pro konverzi do GTFS je možné použít *jrutil* (Koňářik, 2020). V tomto programu nejsou vyřešeny převody datumových omezení a neumožňuje práci s historickými daty. Naopak velmi zajímavá je vyvíjená funkcionalita *jrutil* pro získání poloh zastávek z Open Street Map na základě fuzzy matching. Tento program v minulosti vytvářel každý den aktuální české jízdní řády ve formátu GTFS. V současné době (březen 2020) je ale tato funkce mimo provoz.

Program jdf2gtfs

Po vyhodnocení možnosti použití nebo úpravy stávajících řešení bylo rozhodnuto o vytvoření nového programu *jdf2gtfs* na míru potřebám autora a datového a analytického centra koordinátora veřejné dopravy v Moravskoslezském kraji KODIS. Celý program je napsán v jazyce Python a využívá databázi PostgreSQL a PostGIS. Pro přístup k databázi se používá balíček SQLAlchemy, který umožňuje objektový přístup z Pythonu k databázovým entitám a datům.

Vstupy do programu jdf2gtfs

Vstupem do *jdf2gtfs* je ZIP archiv s jízdním řádem ve formátu JDF. Podporováno je také načtení více jízdních řádů najednou, které jsou zabaleny v dalším zip archivu. Tato funkcionalita je vhodná ke zpracování jízdních řádů z CIS JŘ, které jsou uloženy podle schématu jedna linka = jeden zip archiv a následně všechny ZIP archivy znovu zabaleny do souboru JDF.zip.

Import jízdních řádů do databáze PostgreSQL

Po načtení jednotlivých souborů probíhá import do databáze s kontrolou duplicit dopravců, zastávek, linek a spojů. Dle zdroje JDF je možné zvolit jako identifikátor zastávky název (výchozí) nebo její CIS číslo (pokud je k dispozici). Jelikož jsou stále v platnosti JDF verze 1.9, 1.10 i 1.11, každá s jinou strukturou souborů a polí, bylo nutné vytvořit tři verze importu dat.

Zpracování informací o provozních dnech a datumových výjimkách

Největším rozdílem mezi JDF a GTFS je způsob zápisu a uložení provozních dnů a datumových výjimek. Ve formátu GTFS je možné pro tento účel použít dva soubory - *calendar.txt*, kde se určí v který den týdne spoj jezdí a *calendar_dates.txt*, kde se určí, v které dny spoj výjimečně jede nebo nejede. Druhou možností je použití pouze souboru *calendar_dates.txt*, kde se vypíší všechny dny, kdy spoj jede.

Formát JDF oproti tomu určuje provozní dny spoje v souboru *Pevnykod.txt*. V něm je možné určit jednotlivé dny v týdnu, nebo pracovní dny, nebo neděle a svátky. Druhým souborem je *Caskody.txt*, kde je možné zapisovat provozní dny a výjimky těmito způsoby: jede, jede také, jede jen, nejede, jede jen v lichých týdnech, jede jen v sudých týdnech, jede jen v lichých týdnech od...do..., jede jen v sudých týdnech od...do... Nejprve se tedy české pracovní dny a svátky převedou pomocí balíčku *pandas timeseries* do konkrétních datumů. Poté proběhne zpracování jednotlivých výjimek. Při použití časových nebo časových a pevných kódů v JDF je většinou výsledkem v GTFS použití pouze souboru *calendar_dates.txt*. Při používání pevných kódů jsou výsledkem v GTFS dva soubory *calendar.txt* a *calendar_dates.txt*.

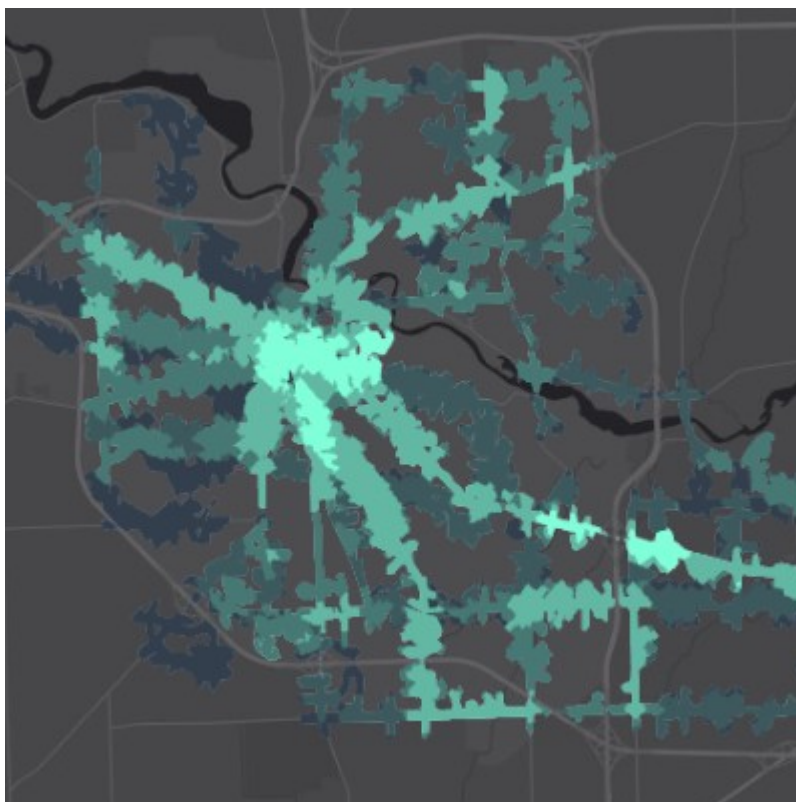
Zpracování prostorových informací

Jelikož formát JDF nepracuje s prostorovou informací, je nutné polohové informace získat z jiných zdrojů. Ve formátu GTFS jsou souřadnice jednotlivých zastávek povinnou součástí souboru *stops.txt*. Tyto informace je možné do programu nahrát z CSV souborů obsahujících název nebo CIS číslo zastávky a její polohu. Další, nepovinnou, prostorovou informací ve formátu GTFS je průběh (linie) jednotlivých autobusových spojů.

Po úspěšném zpracování a import dat je možné pomocí SQL dotazů vygenerovat všechny potřebné GTFS soubory. Pro analytické potřeby je ale důležité znát také historii změn jízdních řádů, což JDF ani GTFS neumožňuje. V těchto formátech je vždy uložen jen jeden aktuálně platný jízdní řád. Proto je možné také vygenerování CSV souboru, který je možné však dobře použít pro analýzy v software jako R nebo SAS, přestože původně vznikl pro interní potřeby KODIS.

PŘÍKLAD POUŽITÍ GTFS V PROSTOROVÝCH ÚLOHÁCH

Webový katalog TransitWiki eviduje okolo 80 aplikací, které pracují s GTFS soubory (TransitWiki, 2020). Kromě těchto povětšinou jednoúčelových aplikací existuje například extenze do ArcGIS s názvem ArcGIS Transit Solution. Ta umožňuje tvorbu a zobrazení prostorových součástí GTFS. Dále podporuje import GTFS do network dataset a následné síťové analýzy. Další možností je tvorba liniového kartogramu intenzit provozu veřejné hromadné dopravy – BetterBusBuffers (Obrázek 1). Některé z těchto funkcí jsou implementovány v základu ArcGIS Pro Network Analyst (Esri, 2020). Podpora GTFS je skrze plug-in MMQGIS a Network obsažena i v QGIS.



Obr. 1. Liniový kartogram vytvořený nad GTFS daty pomocí BetterBusBuffers (zdroj: esri.github.io)

ZÁVĚR

Pro potřeby analytického zpracování jízdních řádů českých autobusových linek byl vytvořen nový program *jdf2gtfs*. Jeho vstupem jsou jízdní řády ve formátu JDF. Konverze probíhá ve třech krocích: import JDF souborů, zpracování informací a import do PostgreSQL databáze a tvorba GTFS souborů pomocí SQL dotazu. Je také možné vytvořit jízdní řád ve formátu CSV, který zachycuje současně i historický jízdní řád. Program *jdf2gtfs* je volně k dispozici na platformě GitHub (<https://github.com/masopust/jdf2gtfs>).

PODĚKOVÁNÍ

Děkuji společnosti Koordinátor ODIS s.r.o. (28. října 3388/111, 702 00 Ostrava – Moravská Ostrava) za poskytnutí originálních JDF jízdních řádů pro Moravskoslezský kraj a za poskytnutí prostorové databáze zastávek. Článek byl podpořen v rámci projektu Výzkum a aplikace metod geoinformatiky pro řešení prostorových jevů reálného světa (IGA_PrF_2020_027) za podpory interní grantové agentury Univerzity Palackého v Olomouci.

LITERATURA

122/2014 Sb. (2014) Vyhláška o jízdních řádech veřejné linkové dopravy. Ministerstvo dopravy, Praha.

11/2014-190-CIS/6 (2014) Metodický pokyn č. 5 k organizaci celostátního informačního systému o jízdních řádech. Ministerstvo dopravy, Praha.

MobilityData. <https://gtfs.org/reference/static>, General Transit Feed Specification Reference, 6. 3. 2020

Zajíčková, L. (2018) Geoinformační přístup správy prostorových dat o veřejné hromadné dopravě na úrovni kraje, Univerzita Palackého v Olomouci, Olomouc

Papež, R. <https://portal.radekpapez.cz/>, Aplikace strojového čtení souborů JDF, 6. 3. 2020

Koňářík, D. <https://gitlab.com/dvdkon/jrutil>, jrutil, 6. 3. 2020

TransitWiki. https://www.transitwiki.org/TransitWiki/index.php/Category:GTFS-consuming_applications, GTFS-consuming applications, 6. 3. 2020

Esri. <https://esri.github.io/public-transit-tools/>, Using GTFS Data in ArcGIS, 6. 3. 2020

Masopust J., <https://github.com/masopust/jdf2gtfs/>, 6. 3. 2020