

**Hledáme jízdni řády  
a v jízdnicích řádech**

# O mně

Student MFF UK, autor [JrUtil](#)

Předchozí dopravní přednášky:

- [Jak nám jezdí vlaky?](#)
- [CIS JŘ: Minulost, současnost a budoucnost jízdnicích řádů](#)

Provozují [fórum s dopravně-datovou tematikou](#).

# Části přednášky

- Zdroje dat pro ČR
- Algoritmy pro hledání spojení

# Typy dat

- Jízdní řády
- Geodata – polohy zastávek, trasy linek
- Real-time polohy a zpoždění vozů
- Data historického provozu

# Data od dopravců/IDS

Jako otevřená data poskytuje JŘ:

- [PID](#) (včetně realtime dat)
- [IDS JMK skrz Brno](#) (včetně realtime dat)
- [DPMLJ](#)
- [DPMO](#) (v GTFS i JDF)

Typicky v GTFS, mezinárodním standardu od Googlu.  
Ten bohužel neumí pokrýt ani všechna specifika JŘ:

- Jen jedna zóna na zastávku
- Neumí popsat zakázané cesty (v rámci města na dálkové lince)
- + další nekritická, ale užitečná, data

Něco pokryjí neoficiální rozšíření, ta ale většinou nepodporuje SW.

Vede k nepěkným datům od dopravců.

# Data z IDOSu

Musím přiznat, CHAPS má bohužel zatím nejlepší data.

Můžeme je obšlehnout! (Můžeme?)

Možná je to dokonce i legální!

*Zeptejte se svého právníka, jestli je využívání cizí databáze vhodné i pro vás. Pozor na autorské právo a zvláštní právo na databáze.*

Jak na to?

- Binární formát IDOSu pro Windows
- Scrapování API CRWS

# Parsování dat IDOSu (.tt)

Několik projektů, některé i zveřejněné:

- [timetab](#)
- Můj nedopsaný popis formátu v [Kaitai Struct](#)
- Slyšel jsem o dalších



# Celostátní informační systém o jízdních řádech

Zákonem zavedený informační systém, který obsahuje všechny jízdni řády veřejné dopravy v ČR.

V provozu od r. 1998, posledních pár let i strojově čitelné exporty.

Teoreticky ideální zdroj JŘ (zákonná povinnost), prakticky má mouchy:

- Nevalidní exporty JDF linkové dopravy
- Místy chybějící výlukové JŘ
- Obsahuje i pár linek, které jsou už mimo provoz

Linkové JŘ ve formátu [JDF](#) (až na záměrné chyby):  
Standard na základě CSV, „jízdni řády podle vyhlášky, ale digitálně“.

Železniční JŘ ve formátu [DJŘ CIS](#):

Export z IS SŽ, XML formát na bázi mezinárodních norem.  
Skoro „kompletní“, obsahuje i třeba informace o IDS, značky pro uživatele („je řazen jídelní vůz“) a čísla kolejí ve stanicích.

Nově také exporty v NeTExu, bohužel nepoužitelné:  
Nejednoznačná ID, nepoužívají verzování.

# Zveřejňování dat z CIS JŘ

Data jsou dostupná až cca od roku 2015, do té doby byla k dispozici jen pro úřady nebo komerčně.

Spol. CHAPS provozovala CIS JŘ zdarma, výměnou za to měla výhradní přístup k datům.

2011 začal Seznam bojovat za zveřejnění dat přes InfZ.

Soudní spory vyšly kladně, Ministerstvo mezitím přidalo povinnost zveřejňovat do nové vyhlášky.

CHAPS tvrdí, že nyní chybějící údaje jsou jejich interní.

# NeTEx

CEN norma pro JŘ a vše okolo.

Velký nekonkrétní XML moloch, „standard na výrobu standardů“.

EPIP – „evropský profil“, konkretizace struktury

Podle [nařízení](#) Evropské komise mají být od letošního prosince dostupné JŘ v NeTExu.

Některé státy už JŘ zveřejňují, v ČR máme nepoužitelné konverze, jinde jen částečná data.

# Materiály k NeTeXu

- Samotný standard je placený, ale... [1](#) [2](#) [3](#) [4](#)  
Tisíce stránek, nedoporučuji číst od začátku do konce.
- [Oficiální vysvětlující dokumenty](#)
- [Oficiální XML schéma](#) a [XML schéma EPIP](#)  
Hodí se použitelná vizualizace, například [moje](#)
- [Oficiální příklady](#)

# Informační systém veřejné dopravy

Nový IS Ministerstva dopravy

Má nahradit/do sebe asimilovat CIS JŘ.

Rok dokončení očekávaný 2026, nejpozději 2027.

Má vyřešit naprosto všechny problémy, dle specifikace...

# „Opravování“ zveřejněných JDF

Ve veřejných exportech (záměrně) chybí důležité údaje: okres a stát u zastávek.

Zásadní problém: Nejednoznačné zastávky

Musíme vytvořit zastávkám unikátní identity a najít je pro každou linku zvlášť.

Identita: Název zastávky + okres + stát

Základní idea: Najdeme v několika průchodech jistá přiřazení názvů zastávek k externím datům poloh/okresů.

- Nemáme 100% pokrytí všech názvů zastávek.  
Tedy si nemůžeme být jisti, že název je unikátní.
  - Různé varianty názvů (čárky, zkratky, někdy i synonyma).
  - Není veřejný dataset zahraničních zastávek podle CIS JŘ
- 
- + Obstojná (polo)veřejná data od většiny krajů/IDS.
  - + Názvy začínají *mají začínat* jménem obce.
  - + V JDF jsou názvy (skoro) jednotné



## „MHD názvy“

V JDF linek MHD mají zastávky někdy „MHD názvy“:  
Chybí název obce, celý název je v jednom sloupečku.

Ještě méně unikátní než plné názvy  
*Mírové náměstí, Nádraží, ...*

Dva způsoby doplnění města:

- Nalezení prefixů k názvům ze seznamu všech zastávek
- Hledání v názvu linky  
„MHD Pardubice linka č.2: ... - ...“ nebo „Zlín,... - Zlín,...“

# Zdroje sekundárních dat

- Polohy zastávek:
  - [Od krajů, měst a IDS](#)
  - [OSM](#)
- Hranice obcí a okresů
  - RÚIAN od ČÚZK
- Polohy obcí mimo ČR
  - OSMNames.org

V repozitáři JrUtil nebo separátně.

# Nepřesnosti názvů, synonyma

Shodu názvů hledám podle Lucene.NET.

Většinou беру pouze data s 100% shodou (bijekce slov).

Řešení zkratk: manuálně napsaný soubor synonym:

nám., náměstí

rozc., rozcestí

odb., odbočka

žel., železniční

nádr., st., nádraží, stanice, stanoviště

...

# Popis algoritmu

- Najdi k zastávkám přesné a unikátní shody podle názvu
- Normalizuj „MHD názvy“:
  - Pro každou zastávku najdi všechny prefixy dávající známý název.
  - Městu podle názvu linky přidej body.
  - Přidej prefix všem zastávkám, kde prefix+název dává známou zastávku, pokud to zvýší počet shod.
- Najdi přesné a unikátní shody podle názvu obce

- Pro každou variantu linky:
  - Projdi zastávky po pořadí a doplň sekundární shody, t.j. omez se na shody s polohami v adekvátní vzdálenosti od předchozí zastávky.
- Doplň okres a stát pro zastávky dostatečně daleko od hranic.

# Kontrola výsledků

Ověření, že každá zastávka má okres/stát.

Ověření vzdáleností zastávek na lince.

Podle kilometráže z JŘ.

Často nefunguje pro mezinárodní spoje (špatné údaje).

# Algoritmy pro hledání spojení

Vstup: jízdní řád, s přestupy nebo polohami zastávek

Výstup: několik nejlepších spojení v konkrétní dobu

Která spojení jsou nejlepší?

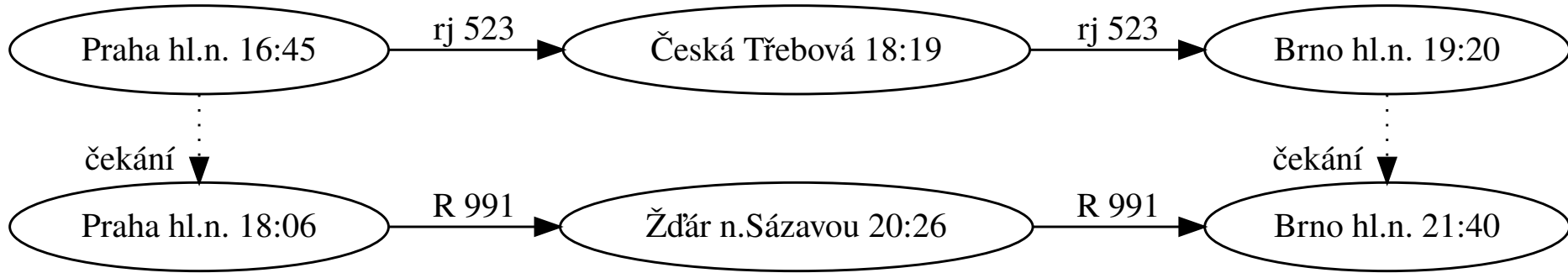
Nejrychlejší, nejméně přestupů, nejlevnější, nejmenší šance na selhání přestupu...

Pareto-optimální spojení: oproti každému jinému má alespoň jedno kritérium lepší.

# Grafová reprezentace JŘ

„Time-expanded“ - pro každou zastávku, každý čas separátní vrchol.  
Velký počet vrcholů, řídký graf

Reálně i víc než níže: zohlednění výstupu/nástupu.





Kompaktnější reprezentace: time-dependent graph

Graf reprezentuje jen síť, bez informací o čase.

Algoritmus si pamatuje k vrcholu vždy čas návštěvy.

Váhy hran jsou funkce hrany a počátečního času.

Nevýhoda: není to standardní vážený graf.

Dijkstrův alg. nebo  $A^*$  na time-expanded grafu funguje, ale...

- Je pomalý oproti algoritmům přímo pro VHD
- Optimalizuje jen čas

# Transfer Patterns

Algoritmus [z roku 2010](#)

Idea: spojení dvou míst mají jen pár optimálních míst přestupu („transfer patterns“).

Velmi rychlý algoritmus, základní verze předpočítá transfer patterns pro každou dvojici zastávek.

[Různá vylepšení](#) sníží cenu předpočítání na desítky hodin.

# RAPTOR

Algoritmus [z roku 2012](#)

Hledá rychle pareto-optimální cesty s mírným předzpracováním.

Používaný většinou dnešního SW: [OTP2](#), [Navitia](#), [MOTIS](#), [r5](#), [rrrr](#)...



- Rozděl spoje podle cesty – sekvence zastávek.
- Nastav  $\text{čas}[\# \text{přestupů}, \text{zastávka}] \leftarrow \infty$ ,  
 $\text{čas}[0, \text{výchozí zastávka}] \leftarrow \text{odjezd}$ .
- Pro  $i$  od 1 do max #přestupů:
  - Pro zastávku  $z$  t., ž.  $\text{čas}[i - 1, z] < \infty$ ,  
 pro  $c$  cestu a první její spoj  $s$  odjíždějící z  $z$  později než  
 $\text{čas}[i - 1, z]$ ,  
 pro každou následující zastávku  $z'$  spoje  $s$ :
    - $\text{čas}[i, z'] \leftarrow \min(\text{čas}[i, z'], \text{příjezd } s \text{ do } z')$
  - Pro  $z', t$  zastávku dosažitelnou z  $z$  a čas na přestup  $t$ ,
    - $\text{čas}[i, z'] \leftarrow \min(\text{čas}[i, z'], \text{čas}[i, z] + t)$

# Další algoritmy

- Connection Scan Algorithm, [z roku 2017](#) (nebo 2013?)
- Trip-Based Public Transit Routing, [z roku 2015](#)

# Úkoly pro vás

Otevřená řešení zde mají potenciál posunout celé odvětví, dobrá řešení jsou ale nad schopnosti jednoho člověka.

- Sběr dat: žádosti na stát/kraj/IDS/města/dopravce, dohody se soukromými subjekty.  
Jízdní řády, historické skutečné jízdy, řazení...
- Programování: Vylepšování serverových vyhledávačů, lokální vyhledávače pro malá zařízení, frontendy, analýzy dat...
- Průběžná kontrola dat, není bohužel možné plně spolehnout na dodavatele.