

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Bakalářská práce

Analýza veřejně dostupných dat o silniční dopravě

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni, 2019

.....

vlastnoruční podpis

Abstrakt

Bakalářská práce se zabývá analýzou veřejně dostupných dat o silniční dopravě. V první až třetí části práce je objasněna teorie týkající se simulace silniční dopravy. Tyto poznatky jsou dále uplatněny v analytické a praktické části práce. Ve čtvrté a páté části jsou analyzovány zdroje vstupních dat se zaměřením na data z dopravního portálu Plzeňského kraje. V šesté části je na základě provedené analýzy navrženo několik způsobů transformace veřejně dostupných dat do podoby uplatnitelné v simulaci silniční dopravy. Sedmá část práce prakticky demonstruje navržené transformace na množině testovacích dat. V poslední části je zhodnocena navržená metodika a způsoby transformace.

Abstract

This thesis deals with the analysis of publicly available data of road traffic. In the first to third part of this thesis the underlying theory of traffic simulation is explained. These findings are then used in analytical and practical parts of the thesis. The fourth and fifth part analyses possible sources for input traffic data with main focus being traffic portal of Pilsen Region. In the sixth part, based on the analysis, a number of transformations of publicly available traffic data are proposed. In seventh part practical demonstrations of proposed transformations are shown on a test data set. Last part of this thesis evaluates used methodology and transformations.

Obsah

1	ÚVOD.....	4
2	ZÁKLADY POČÍTAČOVÉ SIMULACE.....	5
2.1	ÚČEL POČÍTAČOVÉ SIMULACE.....	5
2.1.1	<i>Analytická simulace.....</i>	5
2.1.2	<i>Virtuální prostředí.....</i>	6
2.2	RYCHLOST BĚHU SIMULACE.....	6
2.2.1	<i>V reálném čase.....</i>	7
2.2.2	<i>V upraveném reálném čase.....</i>	7
2.2.3	<i>Co nejrychleji.....</i>	7
2.3	DĚLENÍ SIMULACE DLE POJETÍ BĚHU ČASU.....	8
2.3.1	<i>Spojité simulace.....</i>	8
2.3.2	<i>Diskrétní simulace.....</i>	8
2.3.3	<i>Hybridní pojetí simulace.....</i>	9
3	SIMULACE SILNIČNÍ DOPRAVY	10
3.1	DĚLENÍ DLE ÚROVNĚ DETAILU	10
3.1.1	<i>Makroskopická simulace</i>	11
3.1.2	<i>Mikroskopická simulace</i>	11
3.1.3	<i>Nanoskopická simulace</i>	11
3.1.4	<i>Mezoskopická simulace</i>	12
3.2	MĚŘENÍ ČASU V SIMULACI SILNIČNÍ DOPRAVY	12
4	VSTUPNÍ DATA PRO SIMULACI SILNIČNÍ DOPRAVY	13
4.1	STATICKÁ DATA.....	13
4.1.1	<i>Dopravní mapy.....</i>	13
4.1.2	<i>Dálkový průzkum země.....</i>	14
4.1.3	<i>GIS</i>	14
4.2	DYNAMICKÁ DATA.....	14
4.2.1	<i>Dotazníková metoda.....</i>	14
4.2.2	<i>Dobrovolníci.....</i>	15
4.2.3	<i>Automatizované systémy.....</i>	15
5	VEŘEJNÉ ZDROJE DAT O DOPRAVĚ	16
5.1	ZDROJE STATICKÝCH DAT.....	16
5.1.1	<i>Projekt OpenStreetMap.....</i>	16

5.1.2	<i>Mapy vycházející z OSM</i>	17
5.1.3	<i>Jiné mapové zdroje</i>	18
5.1.4	<i>Nemapové zdroje statických dat</i>	19
5.2	ZDROJE DYNAMICKÝCH DAT.....	20
5.2.1	<i>Kamerový a radarový systém plzeňského kraje</i>	21
5.2.2	<i>Ostatní podobné portály</i>	23
6	DATA Z DOPRAVNÍHO PORTÁLU PLZEŇSKÉHO KRAJE	24
6.1	FORMÁT ULOŽENÍ.....	24
6.2	STRUKTURA A OBSAH ARCHIVŮ.....	25
6.3	OBSAH A VÝZNAM CSV SOUBORŮ.....	26
6.3.1	<i>Formát CSV</i>	26
6.3.2	<i>Locations</i>	27
6.3.3	<i>Dopravní data</i>	28
7	TRANSFORMACE DAT	33
7.1	VÝSLEDNÝ FORMÁT.....	33
7.1.1	<i>Pravidla výsledného CSV formátu</i>	33
7.2	OČIŠTĚNÍ DAT.....	33
7.2.1	<i>Způsob porřízení záznamu</i>	34
7.2.2	<i>Umístění</i>	35
7.2.3	<i>Čas</i>	35
7.2.4	<i>Rychlost</i>	36
7.2.5	<i>Typologie vozidel</i>	36
7.3	AGREGACE DAT.....	37
7.3.1	<i>Bez použití agregace</i>	37
7.3.2	<i>Agregace podle času</i>	38
7.3.3	<i>Agregace podle místa</i>	39
7.3.4	<i>Agregace typů vozidel</i>	40
7.4	ATRIBUTY VÝSLEDNÉHO FORMÁTU.....	40
7.4.1	<i>Identifikace místa</i>	41
7.4.2	<i>Časové zařazení</i>	41
7.4.3	<i>Počet vozidel</i>	42
7.4.4	<i>Rychlost vozidel</i>	42
7.4.5	<i>Typologie vozidel</i>	43
8	DEMONSTRACE POPSANÝCH TRANSFORMACÍ	45
8.1	OČIŠTĚNÍ DAT.....	45
8.2	VYUŽITÍ ČASOVÉ AGREGACE.....	47
8.2.1	<i>Časová agregace s počtem průjezdů</i>	47
8.2.2	<i>Časová agregace s průměrnou rychlostí</i>	50

8.2.3	<i>Časová agregace s typologií vozidel</i>	52
8.2.4	<i>Časová agregace dělená podle typů záznamu</i>	53
8.3	VYUŽITÍ POMĚROVÝCH UKAZATELŮ	55
8.3.1	<i>Časová agregace s poměry typů vozidel</i>	55
8.3.2	<i>Časová agregace s přepočtenými poměry vozidel</i>	56
9	ZHODNOCENÍ VÝSLEDKŮ	57
9.1	ZHODNOCENÍ ANALÝZY DAT.....	57
9.1.1	<i>Nefunkční stanice</i>	57
9.1.2	<i>Nefunkční měřicí zařízení</i>	57
9.1.3	<i>Změna podoby dat</i>	58
9.1.4	<i>Chybějící informace</i>	59
9.1.5	<i>Možný vícenásobný záznam</i>	59
9.2	ZHODNOCENÍ UŽITÝCH TRANSFORMACÍ	60
9.2.1	<i>Informační hodnota nových záznamů</i>	60
9.2.2	<i>Ztrátovost při agregaci</i>	60
9.2.3	<i>Vyhodnocení složitosti algoritmu</i>	61
10	ZÁVĚR	63

1 Úvod

Silniční doprava je nedílnou a důležitou součástí moderní společnosti. Na globální úrovni je faktorem ovlivňujícím mimo jiné ekonomiku, spotřebu energie a životní prostředí. Z toho důvodu je třeba v případě zásahu do dopravní sítě postupovat rychle a zároveň dbát na to, aby provedené změny byly co možná nejefektivnější. V takových případech je důležité umět s co největší přesností předpovědět, jak daný zásah dopravní situaci v daném regionu ovlivní a vyhnout se případným nákladům na nápravu. Právě k tomuto účelu lze jako nástroj využít počítačovou simulaci silniční dopravy.

Pro efektivní fungování takové simulace je však potřeba využít vhodných vstupních dat, která se zdají být díky stále rozšířenějším moderním technologiím lépe dostupná. Využití nejrůznějších automatizovaných záznamových (např. radarových) systémů zejména ve městech se stává běžnou praxí. Data z takových systémů pak bývají archivována v databázích a v některých případech pak ve formě volně dostupných dat k dispozici online ke stažení.

Tato práce si dává za cíl otevřená data z jednoho takového zdroje analyzovat a popsat jejich podobu. Dále na základě analýzy navrhuje možnou transformaci těchto dat z jejich původní podoby do podoby, kdy budou využitelná v různých typech simulace silniční dopravy. Tyto transformace jsou následně demonstrovány na množině testovacích dat. V závěru práce jsou navržené transformace zhodnoceny a je zmíněn další možný postup při práci se zde řešenými daty, nebo daty ze zdrojů podobného charakteru.

2 Základy počítačové simulace

Na začátek je třeba říci, co je to *simulace*. Definice tohoto termínu je hned několik a liší se zejména podle toho, v jakém kontextu a odvětví se o simulaci mluví. V některých případech se definice soustředí na její účel, v jiných na to, jakým způsobem samotná simulace probíhá.

Lze říci, že účelem simulace je reprodukovat chování systému, který je pro tuto potřebu vyjádřen ve formě dostatečně přesného fyzického, logického, matematického, nebo teoretického *modelu*.

Za *počítačovou simulaci* pak lze označit napodobení chování skutečného systému, který je pro tuto potřebu reprezentován matematickým modelem dostatečně zjednodušeným na to, aby mohl být zpracován na výpočetním zařízení, kde má simulace probíhat [1].

V následujících bodech je popsáno několik způsobů dělení počítačové simulace a jsou vysvětleny základní termíny z této oblasti.

2.1 Účel počítačové simulace

Podle účelu lze počítačové simulace rozdělit na dva základní přístupy. Prvním z nich je *analytická simulace* a druhým z nich je *virtuální prostředí*. Tyto dva typy simulace, řeší rozdílné problémy rozdílnými způsoby a liší se v několika základních bodech, kterými jsou cíl simulace, rychlost vykonávání simulace, interakce s lidským faktorem a vztahy na sebe navazujících událostí [2].

2.1.1 Analytická simulace

Účelem analytické simulace je nashromáždění detailních kvantitativních dat komplexního systému pro další využití. Je stěžejní, aby model popisující danou situaci byl co nejvěrnější kopií svého skutečného protějšku, alespoň z hlediska dat, která se snažíme simulací získat. To nám zajistí relevanci získaných dat. Nicméně, i v takové situaci je často vhodné až nutné nechat simulaci proběhnout několikrát, abychom dosáhli námi požadované přesnosti (více [4]). Z tohoto důvodu se často u analytické simulace přistupuje k vykonávání „*co nejrychleji*“ (podrobněji vysvětleno v Kap. 2.2), kdy je člověk ve valné většině případů pouze pasivním pozorovatelem a do samotného průběhu simulace zasahuje jen minimálně, zpravidla jen při jejím spouštění. Dalším

důležitým faktorem analytické simulace je zajištění toho, aby byly vzniklé události vykonány přesně v takovém pořadí, jako by proběhly ve skutečném světě [2].

2.1.2 Virtuální prostředí

Tento přístup k počítačové simulaci je o něco mladší a rozšířil se zejména ve dvou odvětvích. Prvním odvětvím je vojenství, kde je virtuální prostředí využíváno pro výcvik jednotek a tudíž do simulace přímo zasahuje lidský faktor, ale například také pro testování systémů a zařízení raketové obrany, kde simulace reaguje na vstupy právě z takových zařízení a člověk není přímým účastníkem. Druhým a stále se rozšiřujícím odvětvím jsou pak počítačové hry, kde je uživatel zpravidla reprezentován pomocí tzv. „avataru“ přímo v herním prostředí. Avatar je v tomto případě personalizovaným zosobněním uživatele s ohledem na pravidla, omezení a podobu konkrétního prostředí, který mu umožňuje v rámci simulace interagovat s okolím a ostatními účastníky simulace [2].

V těchto případech je tedy účelem vytvoření prostředí, které vypadá co možná nejpřesněji vzhledem k předloze a které uživateli dává pocit, jako by byl přímou součástí simulace. Důležitou věcí je tedy také to, aby běh virtuálního prostředí byl co možná nejplynulejší a zároveň svým tempem odpovídal skutečnému světu. V opačných případech by se mohlo konání v takovém prostředí jevit nepřirozeně, např. přistání s letadlem by se při pomalejším běhu času jevilo snazší. Pro vztahy různých událostí zde, na rozdíl od analytické simulace, platí, že v simulaci nemusí proběhnout přesně ve stejném pořadí, případně přesně ve stejnou chvíli, ale uživatel nesmí být schopen tuto nepřesnost rozpoznat [2].

2.2 Rychlost běhu simulace

V počítačové simulaci se vyskytuje několik různých pohledů na čas. V první řadě se jedná o *fyzický čas*, který označuje časové období, ve kterém simulace probíhá, např. simulace hustoty silničního provozu v období od 1. do 15. ledna roku 2017. Dalším termínem je *simulační čas*, který představuje abstrakci fyzického času do podoby, ve které je v simulaci uchován, např. jednotlivé dny z předchozího případu jsou uloženy jako individuální jednotky s přesností na dvě desetinná místa, čas v simulaci tedy začíná na hodnotě 1.00 a končí na hodnotě 15.00. Posledním zde zmíněným termínem je pak *skutečný čas* (wallclock time), který představuje dobu, po jakou daná simulace běžela

ve skutečnosti, např. ve skutečnosti simulace probíhala 30. prosince roku 2016 od 13. do 15. hodin [2].

Podle vztahu těchto hodnot pak lze rozdělit počítačovou simulaci na několik základních kategorií, kterými jsou *simulace v reálném čase*, *simulace v upraveném reálném čase* a *simulace co nejrychleji*.

2.2.1 V reálném čase

Tento způsob simulace je založen na tom, že simulační čas ubíhá stejnou rychlostí, jako skutečný čas, což je žádoucí zejména u virtuálních prostředí, kde by v případě pomalejšího běhu simulačního času oproti skutečnému času působila simulace loudavě a neresponzivně vůči uživateli. Kupříkladu u vojenských cvičení by se splnění úkolů zdálo snazší, než je ve skutečnosti. V případě opačném, kdy by čas simulace běžel rychleji, než čas skutečný, by se naopak úkol mohl zdát těžší a pohyb v takovém prostředí matoucí.

Simulacím probíhajícím v reálném čase se souhrnně říká *realtimové simulace* [2].

2.2.2 V upraveném reálném čase

Původní anglický název „scaled realtime simulation“ vyjadřuje podstatu tohoto způsobu běhu simulace o něco lépe. Jedná se o způsob podobný běhu v reálném čase s tím rozdílem, že je simulační čas oproti skutečnému času upraven (zrychlen, nebo zpomalen) o nějakou konstantní modifikační hodnotu – „škálován“. Například jedna vteřina skutečného času může znamenat uplynutí tří vteřin v simulačním čase. Tento přístup není nepodobný kupříkladu rychlému přetáčení VHS kazety. V praxi se používá ke zrychlení nezajímavých nebo nepodstatných částí simulace ve virtuálním prostředí, nebo naopak ke zpomalení důležitých sekvencí pro podrobnější analýzu nastalé situace [2].

2.2.3 Co nejrychleji

Je způsob uplatňovaný zejména u analytických simulací a vyznačuje se tím, že jediný vztah mezi simulačním časem a skutečným časem je ten, že skutečný čas nám řekne, jak dlouho jsme museli čekat na dokončení simulace. Zde se simulací prochází nejrychlejším možným způsobem a jednotlivé části simulace nemají pevně určený časový rámec. Některé mohou proběhnout během několik vteřin, kdežto jiné mohou zabrat desítky minut nebo hodin [2].

Při vhodném určení vztahu mezi simulačním a skutečným časem je pak možné docílit toho, že simulace může běžet při jednom spuštění způsobem v reálném čase a při spuštění jiným způsobem co nejrychleji [2].

2.3 Dělení simulace dle pojetí běhu času

U počítačové simulace lze vztah k pojetí běhu času rozdělit na dvě základní kategorie. První kategorií je *spojitá simulace* a druhou *diskrétní simulace*.

2.3.1 Spojitá simulace

Základem spojitě simulace je zpravidla soustava diferenciálních rovnic, která je schopná vyjádřit stav fyzického systému v jakémkoliv časovém bodě. Ústředním bodem jsou zde vztahy mezi komponentami systému, které zde bývají složité a nelineární a výpočty zahrnují zpracování zpětných vazeb [10]. Ve spojitých simulacích se ve valné většině případů pracuje s agregovanými daty a konkrétní prvky se nepoužívají. Zároveň se nevyužívá náhodných proměnných a lze je tedy označit za deterministické. Výsledky takové simulace jsou jednoznačné a je možné je bez dalšího kontextu a statistického zpracování využít v další práci [3].

Spojitá simulace se využívá například pro předpovídání počasí, dynamiku populací a ekosystémů [3].

2.3.2 Diskrétní simulace

Stavebním kamenem diskrétní simulace je pojetí času jako časových skoků s pevnou délkou. Tento systém se zaměřuje na jednotlivé, konkrétní *entity* a *události*, každá z entit má jednu nebo více unikátních *vlastností*, které ovlivňují její chování v systému. Vztahy mezi entitami v této simulaci jsou jednodušší povahy a lineární. Pro události platí, že vznikají ve chvíli, kdy je některá z entit ovlivněna vykonáváním *aktivity*. Hodnoty v diskrétní simulaci (např. rychlost vozidla, délka fronty), bývají často ovlivněny náhodným prvkem, který dále podporuje individualizaci, a její průběh je v takovém případě stochastický [4].

Diskrétní simulace dále používá *globální proměnné*, které jsou společné a ovlivňují všechny entity, *kalendář událostí*, který obsahuje události, jež mají do systému vstoupit v budoucnu, a *globální čas*, jenž určuje, v jakém bodě se simulace právě nachází. Vzhledem ke stochastické povaze diskrétní simulace je pro získání dostatečně relevantních výsledných dat často nutné vykonat simulaci několikrát (více i s příklady v

[4] a [10]) a výsledná data dále matematicky zpracovat. Příkladem diskrétní simulace může být kupříkladu simulace silniční dopravy, simulace nákupního centra a obecně systémy zahrnující fronty [4].

Pro přechod z jednoho časového bodu do druhého se běžně používají dva způsoby. Prvním z nich je *pevný časový krok*, kde dochází k přepočítání proměnných pro všechny entity, a to i v případě, že nebyly nijak ovlivněny událostí. Druhým ze způsobů je systém *řízený událostmi*, kde dochází k přepočítání pouze těch proměnných, které byly v daném časovém kroku ovlivněny některou z proběhlých událostí. Tento způsob je běžnější a využívá se zejména v případech, kdy události probíhají řídce vzhledem k času [2].

2.3.3 Hybridní pojetí simulace

Stejně jako v mnoha jiných odvětvích i zde je efektivní volbou využít silných stránek obou pojetí simulace pro dosažení kýžených výsledků v co nejkratším čase a v požadované přesnosti. Práce s takovými modely je však nesnadná, neboť vyžaduje hluboké znalosti z obou oborů a k využití tohoto potenciálu je třeba vynaložit neúměrně více úsilí. Více informací v [3].

3 Simulace silniční dopravy

V tomto oddíle jsou definovány důležité pojmy z oblasti simulace silniční dopravy a případná specifika oproti obecnější počítačové simulaci.

Důležitými prvky pro simulaci silniční dopravy, které se mohou lišit podle úrovně detailu, jsou:

Přesné a realistické topologické mapy – v základu určující prostor, po kterém se doprava může pohybovat a to včetně dodatečných detailů jako je typ silnice, počet pruhů a jiné (více v [12]),

Hladké zrychlení a zpomalení – aby chování vozidel v rámci simulace co nejrealističtěji reprezentovalo skutečný pohyb, nelze tyto prvky uvažovat pouze jako jednorázový a okamžitý skok,

Překážky – jsou obecně entity, které dále omezují prostor, po kterém se reprezentovaná vozidla mohou v simulaci pohybovat; lze je rozdělit na *dynamické* (interakce mezi vozidly), nebo *statické entity* (uzavírka). V širším pojetí toho pojmu se jedná také o *informační ruchy*, které omezují možnost pohybu v modelu, způsobené např. použitým komunikačním protokolem,

Generátor a terminátor – jsou místa, ze kterých do modelu vstupují, respektive ve kterých model opouštějí vozidla nebo obdobné entity a prvky, zpravidla se nacházejí na okraji prostoru simulace,

Simulační čas – pomáhá realisticky určit časový kontext, např. pro vyšší hustotu dopravy ve špičkách,

Členění vozidel – rozčleňuje vozidla do jednotlivých kategorií, které určují jejich velikost, maximální rychlost, zrychlení a jiné parametry,

Chování řidiče – obdobně jako v předchozím bodě člení řidiče jednotlivých vozidel do různých kategorií, které mají rozdílné chování a reakce při vyhodnocování situací [11].

3.1 Dělení dle úrovně detailu

S přihlédnutím k úrovni detailu v daném modelu se simulace silniční dopravy dělí na dvě základní skupiny. Těmi jsou *makroskopická simulace* a *mikroskopická simulace*.

Z těchto dvou skupin lze ještě dále vyčlenit *nanoskopickou simulaci*, která navazuje na mikroskopickou a dále ji konkretizuje, a *mezoskopickou simulaci*, která se pohybuje na pomezí makroskopické a mikroskopické [7].

3.1.1 Makroskopická simulace

V makroskopickém pojetí simulace je doprava vyjádřena pouze ve formě agregovaných dat, kde se neuvažují žádná individuální vozidla, a která je popsána funkcí prostoru a času. Typickými sledovanými vlastnostmi v takových modelech jsou pak *průměrná rychlost, plynulost dopravy a koncentrace* v daném bodě [5].

Jedná se o nejstarší přístup, který má své výhody zejména v nejmenší náročnosti na výpočty [6]. Příkladem může být makroskopický model založený na Gas Kinetic Theory [5].

3.1.2 Mikroskopická simulace

Na rozdíl od makroskopické simulace se v mikroskopické simulaci pracuje s jednotlivými vozidly, které se pohybují ve spojitém (např. Car-following model – model následování vozidel [7]), nebo diskrétním (např. Cellular automaton model – model založený na buňkovém automatu [14], [16]) prostoru. V některých případech jsou zde vlastními entitami také řidiči jednotlivých vozidel a uvažuje se jejich chování. Pohyb vozidel je v těchto modelech uvažován buď v 1D prostoru, kdy se vozidlo pohybuje pouze v jednom pruhu, nebo v případě výskytu více pruhů mezi nimi okamžitě „přeskakuje“, nebo ve 2D prostoru, kde je přesun z pruhu do pruhu zohledněn a vyjádřen pohybem po hladké křivce, což umožňuje např. realistické předjíždění v modelu [11]. Mezi vlastnosti vozidel patří například jejich *pozice, individuální rychlost* nebo *snaha zrychlit* a dosáhnout své *optimální rychlosti* [6].

Vzhledem k detailnosti a množství jednotlivých prvků mikroskopické simulace se jedná o výpočetně výrazně náročnější způsob simulace než v případě makroskopické simulace [7].

3.1.3 Nanoskopická simulace

S rostoucí výpočetní silou HW se v některých situacích mikroskopický model začal členit ještě detailněji a tento přechod se dá zařadit do vlastní kategorie nanoskopické simulace. Zvláštní důraz je zde kladen kupříkladu na rozsáhlejší individualizaci vozidel, kde každý typ vozidla odpovídá jinak na pokyn řidiče (např. rychlost reakce, zrychlení,

ochota zatočit), a řidičů, kteří jsou členěni do několika kategorií (např. běžný, opatrný, agresivní), mají vlastní „inteligenci“ a reagují různě na své okolí. Dalším detailem je pak skutečný pohyb vozidla v 2D prostředí, kde nemusí docházet pouze k přesunům z pruhu do pruhu, ale prostor silnice lze uvažovat spojitě a vozidlo se tak může volně pohybovat v rámci svého směru, nebo celé silnice (například v [9]). Nanoskopická simulace je blíže popsána například v [6].

3.1.4 Mezoskopická simulace

Snahou těchto modelů je popsat daný systém s podobným důrazem na detail jako u mikroskopické simulace, ale zároveň se zjednodušením vyhnout zatěžujícím výpočtům pro jednotlivá vozidla s využitím principů makroskopické simulace. Příkladem může být sjednocování vozidel do skupin, kde se uvažuje rychlost skupiny a počet vozidel ve skupině a také zjednodušení vztahů, kdy spolu jednotlivé skupiny interagují pouze v dopravních uzlech (např. křižovatky, vjezdy). Kategorizace konkrétních modelů a principů se pak v mnoha případech liší zdroj od zdroje, podle způsobu definice jednotlivých druhů simulace [8].

3.2 Měření času v simulaci silniční dopravy

V simulaci silniční dopravy je uvažováno diskrétní pojetí času, které průběh simulace rozděluje na kroky pevné časové délky (často 1 sekunda). Tento způsob se dá dále uplatnit využitím metody *pevného časového kroku*, nebo systému *řízeného událostmi*.

V případě využití pevného časového kroku jsou hodnoty propočítávány pro všechny entity při každém kroku simulace. Tento přístup je využíván ve většině případů, které počítají se širší možností interakce mezi jednotlivými vozidly. Ta může probíhat nejen na křižovatkách a jiných významných dopravních uzlech, ale i při pohybu po běžné jednosměrné silnici.

Řízení událostmi lze využít v případě, kdy model zachází se zjednodušenými vztahy mezi entitami (např. k interakci dochází pouze v dopravních uzlech) a nedocházelo by tak ke generování velkého množství událostí, který by tento přístup učinil neefektivním [13].

4 Vstupní data pro simulaci silniční dopravy

Pro věrné fungování simulace silniční dopravy je třeba získat různorodá vstupní data. V nejlepším případě jsou data v již požadovaném formátu dostupná online a je tedy možné jejich přímé využití [21]. V opačném případě je nutné vstupní data filtrovat a případně transformovat, aby bylo jejich využití možné.

K dispozici je množství přístupů k získávání nebo vytvoření těchto dat, které lze navíc efektivně kombinovat pro zaručení dostatečné přesnosti získaných dat. Sběr a transformace těchto dat probíhá s využitím lidského faktoru [19], automatizovanou formou [12], nebo kombinací obojího.

Potřebná data lze dále rozdělit na *statická data* a *dynamická data*, která jsou stručně popsána v následujících podkapitolách.

4.1 Statická data

Cílem sběru statických dat je vytvoření co nejrealističtější dopravní sítě pro konkrétní simulaci silniční dopravy. Využívá se informací, jako je např. typ silnice, který omezuje maximální rychlost vozidla, počet pruhů silnice v obou směrech, komplexní informace o křižovatkách, včetně odbočovacích pruhů, zúžení, nájezdových ramp, ale také informace o semaforech a jejich časování. Dále se do této kategorie řadí informace o dlouhodobých uzavírkách komunikací a omezeních, které budou simulaci ovlivňovat po celou dobu jejího běhu, ale také výskyt případných menších překážek. Jedná se kupříkladu o uzavírku celé komunikace, omezení maximální rychlosti vozidel, nebo omezení provozu uzavřením jednoho, nebo více jízdnic pruhů dopravní komunikace [12].

Statická data lze získat z různých zdrojů. Jednotlivé možnosti jsou stručně popsány v následujících kapitolách.

4.1.1 Dopravní mapy

Jednou z možností je získat data z map, které jsou určeny přímo k popisu dopravní sítě. V případě větších měřítek jsou schopné poskytnout přesné informace o jejím rozložení a o typech silnic, ale podrobnější informace potřebné pro simulaci, jako například popis vjezdů a výjezdů křižovatek, nebo počet jízdnic pruhů často nejsou z těchto zdrojů

snadno dostupné [17]. V takovém případě je nutné využít jiných doplňkových zdrojů a možností.

4.1.2 Dálkový průzkum země

Další možností je dálkový průzkum země. Jedná se o bezkontaktní sběr dat o zemském povrchu, který nejčastěji probíhá pomocí snímkování z letadel nebo satelitů. Předpokladem je sběr dat v dostatečném rozlišení a jasný výhled na prvek, který chceme zaznamenat a transformovat [12], [18].

Tento typ sběru dat má díky pokroku a zvyšující se přesnosti a frekvenci snímání potenciál stát se i zdrojem pro sběr dynamických dat o vozidlech [18].

4.1.3 GIS

Získávání dat z geografického informačního systému (GIS) se zakládá na transformaci dopravních dat v daném modelu GIS z křivek do podoby, jakou využívá daný typ simulace silniční dopravy. Důležitou výhodou je zde také možný výskyt patričních doplňkových dat, což mohou být již zmíněné počty pruhů, typy silnic a popis křižovatek [12]. Tento způsob sběru a transformace dat lze efektivně využít s předchozím bodem.

4.2 Dynamická data

Mezi dynamická data lze zařadit informace o vozidlech, která do silniční sítě modelu vstupují na předem definovaných generačních bodech, nebo která ji opouštějí v terminačních bodech. Základními informacemi pro tyto prvky je například rozložení typů vozidel (např. osobní vozidla, motocykly, autobusy, kamiony a kamionové soupravy), množství vozidel, které projede daným bodem za jednotku času (tj. hustota dopravy) a rychlost jednotlivých vozidel. Mezi dynamická data se rovněž řadí pravděpodobnost odbočení v křižovatkách či celé trasy jednotlivých vozidel [12].

Opět existuje více používaných možností, jak tato data získat. Běžné metody jsou popsány v následujících podkapitolách.

4.2.1 Dotazníková metoda

V případě dotazníkové metody se nejedná o konkrétní počítání průjezdů aut nebo hustoty dopravy, ale o sběr dat od osob, které v dané oblasti žijí a využívají automobilů jako dopravních prostředků. Při dostatečně velkém vzorku je tato metoda schopna poskytnout informace o typech aut, které v daném úseku projíždějí, výskytech důležitých uzlů a pro dopravu zajímavých míst, které mohou následně sloužit jako

generační a terminační body, ale také atraktivních a vytížených dopravních úseků. V neposlední řadě mohou v případě nanoskopického pohledu na simulaci poskytnout i náhled do povah řidičů, což umožní jejich následnou kategorizaci [19].

Tento způsob sběru informací lze pro efektivitu doplnit například dlouhodobým sběrem dat od řidičů s využitím GPS lokátorů pro další modelování chování řidičů [19].

4.2.2 Dobrovolníci

Dalším ze způsobů sběru dynamických dat je využití dobrovolníků pro sběr relevantních informací v daném úseku na strategicky umístěných bodech. Takovými informacemi může být počet a typ aut, které daným místem projely za určitý časový úsek, počet chodců, nebo výskyt uzavírek a jejich dopad. Tuto metodu je možné využít také např. k počítání lidí využívajících MHD a na základě nasbíraných dat provést optimalizaci jednotlivých linek [19].

Tento přístup lze kombinovat s přístupem z následujícího bodu, kdy může být lidský faktor využit pro vyhodnocení kamerového záznamu, případně pro rozsouzení nerozhodných případů, se kterým si software sám neporadí.

4.2.3 Automatizované systémy

V současné době je tento způsob sběru dat na neustálém vzestupu. Ať už se jedná o sběr dat přímo ve městech na důležitých a zajímavých místech z informativního důvodu, z důvodu bezpečnosti dané lokality, nebo o sběr dopravních dat za účelem archivace a dalšího zpracování. V případě kamerových a radarových systémů umožňují pokročilé metody analýzy obrazu využití i těchto dat pro potřeby simulace silniční dopravy. S využitím takových obrazů jako zdrojů lze vlastním postupem získat potřebné informace o dynamických prvcích silniční dopravy. V jiných případech lze využít již zpracovaná data v textové podobě a ty vhodným způsobem vytržít a transformovat [20]. V tomto případě mohou být zdrojem takových dat opět kamerové systémy, ale také některé jiné senzory, kupříkladu indukční smyčky, váhy, nebo detektory založené na přerušení paprsku světla, které jsou schopné shromažďovat data o průjezdech vozidel, v některých případech také o jejich rychlostech [34].

5 Veřejné zdroje dat o dopravě

Tato část práce se zaměřuje na prozkoumání některých veřejně dostupných zdrojů mapových dat a potenciál jejich využití jako statických dat pro modelování dopravní sítě, ale také využití jiných, zejména obrazových, vstupů jako zdroj pro dynamické prvky simulace silniční dopravy.

U statických dat jsou hodnoceny zejména následující klíčové vlastnosti. Schopnost realisticky reprezentovat původní dopravní síť s ohledem na zachování přesných vzdáleností a rozložení dopravní infrastruktury. Dále se jedná o možnost určení počtu pruhů dané komunikace, určení maximální možné rychlosti na úseku, ale také o komplexní analýzu křižovatek, nájezdů a sjezdů a slučovacích pruhů pro jejich realistické zaznamenání. V rámci tohoto bodu jsou také uvažovány možné zdroje dlouhodobých i krátkodobých dopravních uzavírek a omezení.

Pro dynamická data jsou hodnoceny následující vlastnosti. U celkové dopravy se hodnotí počet průjezdů v daném úseku, z čehož lze zjistit hustotu dopravy. U jednotlivých vozidel se hodnotí rychlost průjezdů, typ vozidla a případná možnost kategorizace řidiče na základě jeho chování.

5.1 Zdroje statických dat

Přestože by se na první pohled mohlo zdát, že taková data jsou v současné době již snadno dostupná, není tomu tak vždy. Informace o dopravní síti jsou volně dostupné online v mnoha podobách, avšak v případech, kdy se nejedná o data určená přímo pro simulaci silniční dopravy, většinou postrádají jednu, často i více klíčových informací. Mimo to, že je mnohdy třeba data čerpat z několika zdrojů, aby se dosáhlo jejich úplnosti, je také třeba získaná data transformovat z podoby, ve které jsou poskytována, do podoby, kterou je možné využít jako vstup pro simulaci silniční dopravy.

5.1.1 Projekt OpenStreetMap

Mezi nejjednodušší a zároveň běžně používané potenciální zdroje dat pro dopravní síť patří online mapové aplikace s interaktivními mapami. Základem pro mnoho takových map je projekt Open Street Map¹, dále jen OSM. Jedná se o soubor otevřených mapových dat, která jsou tvořena a spravována rozsáhlou a různorodou komunitou

¹ <https://www.openstreetmap.org>

příspěvatelů z celého světa. Ta si klade za cíl vytvořit co nejpřesnější mapový základ a udržovat ho neustále aktuální a přesný [22].

K těmto datům je možné přistupovat přímo v prostředí prohlížeče, kdy interaktivní okno poskytuje informace o infrastruktuře dopravní sítě, typech silnic a směrech provozu, nebo je možné se registrovat a přihlásit. V tu chvíli je uživatel nejen schopen data v online editoru upravovat, ale také zobrazit dodatečné informace k jednotlivým prvkům. Z užitečných informací se jedná o určení, zda je silnice jednosměrná, nebo obousměrná, jaká je maximální povolená rychlost, doporučená rychlost a jaký je počet pruhů.

Další velmi užitečnou vlastností je možnost exportu vybrané lokality, nebo komunikace do formátu XML. Jedná se o vektorová data, kde dochází na začátku souboru k definování všech uzlů vybrané lokality a jejich umístění, následně jsou s jejich pomocí definovány linie, kupříkladu vodních toků, silnic, cyklotras, vedení vysokého napětí, a plochy, kupříkladu pozemky, územní celky, jezera a rybníky. V daném souboru se tak kromě informací o dopravní síti nachází mnoho redundantních informací. V případě exportu jednotlivých komunikací obsahuje soubor pouze data o dané linii a jeho uzlech, informace o nich už však chybí [22].

Hlavními nevýhodami tohoto zdroje je nutnost tato data filtrovat a transformovat, stejně jako fakt, že není povinné u silnic vyplňovat všechny atributy a tak kupříkladu informace o počtu pruhů velmi často chybí, zejména u obousměrných silnic. Další možnou nevýhodou je otevřenost komunity OSM a tak může, zejména v oblasti editované málo uživateli, dojít k výskytu nepřesných, nebo nepravdivých informací z důvodu případné nedostatečné kontroly.

5.1.2 Mapy vycházející z OSM

Projekt OSM vzhledem ke svým autorským právům a licencím nabízí svá data volně za předpokladu, že budou šířena pod stejným licenčním ujednáním [22]. To vede k jejich rozšířenému užití i na některých českých mapových portálech. V následujících bodech bude diskutováno několik z nich včetně toho, jaké informace navíc jsou schopné nabídnout.

openstreetmap.cz

V tomto případě se jedná o stránky české části komunity OSM, která se zaměřuje zejména na zvýšení kvality mapování na území České republiky. Tento portál nenabízí možnost data přímo editovat, ale nad rámec původního projektu umožňuje zobrazení ortofotomapy vytvořené Českým úřadem zeměměřičským a katastrálním (ČÚZK). Rozlišení těchto snímků umožňuje vizuální kontrolou využitím lidského faktoru určit některé důležité atributy dat, které z předchozích vektorových dat chybějí. Může se jednat o určení počtu pruhů dané komunikace, tvaru křižovatky nebo dopravního uzlu, ale také o zjištění vodorovného dopravního značení [23].

mapy.idnes.cz

Jedná se o mapovou aplikaci provozovanou v rámci serveru idnes.cz, který je ve správě společnosti MAFRA a.s. Tento mapový portál vychází ze základních mapových dat OSM, ze kterých dále využívá pouze základní a turistickou mapu. Tato data navíc není možné z tohoto portálu přímo exportovat pro potřeby dalšího využití. Portál proto nemá pro tuto práci ani simulaci silniční dopravy přílišný potenciál, neboť veškeré nabízené relevantní funkce efektivněji pokrývají jiné zdroje [24].

mapy.cz

Jedná se o mapovou aplikaci provozovanou v rámci serveru seznam.cz společností Seznam.cz a.s. Aplikace vychází ze zdrojových dat OSM, obdobně jako předchozí případ, avšak nabízí širší sortiment map. Za zmínku v tomto případě stojí hned několik verzí ortofotomap z let 2003, 2006, 2012, 2015 a současně aktualizované verze. Snímky pro nejnovější verzi této mapy jsou poskytovány společností TopGIS s.r.o. a v rozlišení 12,5cm/pixel [28] umožňují, obdobně jako v prvním bodě, určit počet pruhů silnice, tvar křižovatky nebo dopravních uzlů a zjistit vodorovné dopravní značení.

Mimo to nabízí mapy.cz také funkci „Panorama“. Jedná se o pozemní snímkování z kamery umístěné na jedoucím automobilu, které zaznamenávají okolí vozidla v rozsahu celých 360°. Tato funkčnost je schopna doplnit informace získané z ortofotomap ve špatně viditelných místech, ale navíc získat informace o svislém dopravním značení [25].

5.1.3 Jiné mapové zdroje

V tomto bodě jsou zmíněny jiné mapové portály, které nevycházejí z projektu OSM, ale používají své vlastní zdroje.

Mapy Google

Jedná se o internetovou, online dostupnou mapovou aplikaci provozovanou společností Google. Tento mapový portál poskytuje a podporuje mnoho funkcí včetně přidávání externích modulů, avšak s ohledem k povaze práce budou zmíněny a zváženy jen některé, relevantní. Aplikace umožňuje interaktivně zobrazit dopravní síť České republiky v podobě vektorové mapy, avšak jakékoliv další informace k silnicím jsou kusé. Dálnice a silnice prvních tříd jsou označeny pouze svým číslem, ostatní komunikace nejsou označeny vůbec, respektive jménem ulice, případně PSČ oblasti, kde se nacházejí. Tato data o dopravní síti navíc není možnost nijak exportovat pro další případnou práci a transformaci. Na stranu druhou Mapy Google umožňují zobrazit mapu z pořízených satelitních snímků v dostatečně vysokém rozlišení umožňujícím vizuální analýzu počtu pruhů silnice, tvaru dopravních uzlů, ale i vodorovného značení. Dále je zde, obdobně jako u mapy.cz, dostupná funkčnost Street View. Jedná se o panoramatické fotografie v okruhu 360° kolem snímajícího vozidla. Tato funkčnost umožňuje další analýzu špatně viditelných míst ze satelitních snímků, případně také kontrolu svislého dopravního značení [26].

Bing Maps

Jedná se o mapovou aplikaci provozovanou společností Microsoft. Na území České republiky umožňují zobrazení základní vektorové mapy, zahrnující i dopravní síť, avšak bez možnosti si jakoukoliv silnici označit a získat o ní dodatečné informace. Rovněž není možné mapy exportovat pro další zpracování a transformaci. Ačkoliv Bing Maps nabízejí možnost zobrazení mapy satelitního snímkování, tyto snímky nejsou často v dostatečně vysokém rozlišení, aby z nich bylo možné získat konzistentní užitečné informace, jako třeba počet pruhů nebo vodorovné dopravní značení. Výhodou tohoto zdroje (vzhledem k datu zpracování práce) je aktuálnost foto snímků. Je možné je tedy využít pro konzultaci s jinými obrazovými zdroji. Dále Bing Maps nabízejí funkci Bird's Eye, která je podobná Panorama od mapy.cz a Street View od Mapy Google, avšak ta není na území České republiky v době zpracování práce dostupná [27].

5.1.4 Nemapové zdroje statických dat

Tento bod se již nezaměřuje na poskytnutí dat o struktuře dopravní sítě, nýbrž na poskytnutí dodatečných informací, které ji mohou v některých místech významně ovlivnit. Může se jednat o významné faktory, jako jsou dlouhodobá omezení dopravy, nebo výskyty uzavírek, které tak mohou ovlivnit chování vozidel v části sledované

oblasti, nebo v jejím celku. Přehlednutí těchto faktorů tak může vést k chybnému vyhodnocení získaných dat. Např. náhlé snížení průjezdů vozidel ve sledovaném bodě nemusí znamenat pokles vozidel v celé oblasti, ale pouze výskyt dopravního omezení na dané komunikaci.

Geoportál plzeňského kraje

Tento portál nabízí v rámci svého GIS několik mapových vrstev, které mají potenciál využití při simulaci silniční dopravy. V rámci projektu „Pořízení a integrace dat pro management regionální silniční sítě PK“ došlo kupříkladu v letech 2014 a 2015 ke sběru dat o svislém dopravním značení na silnicích 2. a 3. třídy. Tato data mohou přímo poskytnout informace kupříkladu o jednosměrných silnicích, případně o rychlostních omezeních v daných úsecích. Jako doplňkový zdroj obdobného charakteru mohou posloužit otevřená data nabízená některými městy, která rovněž obsahují data o dopravním značení. Zároveň to usnadňuje případné porovnání dat z panoramatických kamer na vozidlech, jež byly zmíněny dříve, neboť není nutné procházet komunikace po celé délce, ale pouze potenciálně zajímavé úseky [29].

Uzavírky a omezení

V tomto bodě jsou zvažována data z dopravního portálu ČR². Tento portál obsahuje několik datových vrstev poskytujících mimo jiné informace o stupních provozu dopravy v dálniční síti České republiky. Dále jsou uvedeny informace o událostech omezujících provoz krátkodobě, jako například znečištění vozovky nebo dopravní nehodě. Nakonec tento portál poskytuje informace o uzavírkách a omezeních dlouhodobějšího charakteru, včetně informací o umístění a rozsahu omezení a jeho časovém trvání. Tyto informace mohou být hodnotným zdrojem pro zjištění důvodu nestandardního chování sběrného místa dynamických dat, kupříkladu kamery, s ohledem na možný nárůst provozu vzhledem k uzavírce alternativní komunikace, nebo pokles či úplné zastavení provozu vzhledem k uzavírce vlastní komunikace [28].

5.2 Zdroje dynamických dat

Navzdory hojnému rozšíření a používání informačních technologií a informačních systémů není jejich potenciál pro sběr dat tohoto typu k dalšímu zpracování a využití v oblasti simulace silniční dopravy zdaleka naplněn. Zvýšení efektivity tohoto využití může vést nejen k úspoře časové, ale také finanční.

² <http://www.dopravniinfo.cz>

V tomto bodě je diskutováno několik možných přístupů k získávání takových dat z veřejně dostupných zdrojů. Zejména se jedná o otevřená data o průjezdech vozidel z radarových stanic na území Plzeňského kraje, která jsou zaznamenávána v rámci česko-bavorského projektu Klidné příhraničí³. Dále se jedná o data z obdobných zdrojů, tedy radarových stanic, v rámci ostatních krajů a zhodnocení jejich relevance.

5.2.1 Kamerový a radarový systém plzeňského kraje

Jedná se o webový portál založený na šabloně pro dopravní webové portály firmy GEMOS CZ [30]. Hlavním účelem portálu je poskytovat přehledná, souhrnná statistická data ze zaznamenávajících radarových stanic v rámci plzeňského kraje, které se nacházejí převážně na frekventovaných silnicích na okrajích měst [32].

Povaha výstupu těchto dat se odvíjí od účelu této části projektu, kterým je s pomocí těchto radarových stanic motivovat řidiče k ukázněné jízdě na území měst a vesnic podílejících se na tomto projektu. Soubory s měsíčními statistickými výstupy pro jednotlivé lokality jsou proto, kromě souhrnných dat o průjezdech vozidel po dnech, zaměřeny hlavně na rozdělení řidičů podle rychlosti – zda překročili povolenou rychlost a jaká byla jejich následná tendence, zda pokračovali dál stejnou rychlostí, zrychlili, nebo zpomalili. Tyto měsíční výstupní soubory jsou uloženy ve formátu PDF a využívají jednotný formát dokumentu [35].

Data jsou poskytována pro každou radarovou stanicí zvlášť. Je k nim možno přistupovat z hlavní stránky odkazem *Města a obce*, odkud je uživatel přesměrován na interaktivní mapu, jejíž základ je poskytnut v rámci projektu OSM, kde jsou zvláště jednotlivá města zapojená do projektu. Při kliknutí na lokalitu je uživatel přesměrován na stránky jednotlivých kamer, kde si na mapě v pravé části může vybrat konkrétní kameru, jejíž data chce prozkoumat. Obdobným způsobem se na tuto stránku může uživatel dostat z rozbalovacího menu *Oblasti*, kde je mu zobrazen seznam zapojených měst, odkud je přesměrován. Vlastní informace o radarové stanici jsou poté zobrazeny v levé části stránky [32].

Jedná se o již zmíněné měsíční souhrny, ale také souhrny po dnech, kdy se jedná o graf počtu průjezdů vozidel ve směru k zařízení rozdělený po hodinách. Zde je možné si zvolit libovolné datum od zahájení provozu zařízení, stejně jako je možné se pomocí rozbalovacího seznamu dostat na statistiky jiné radarové stanice ve stejné oblasti,

³ Více informací o projektu zde: <http://www.plzensky-kraj.cz/cs/clanek/klidne-prihranici>

respektive stejné obce. Na předchozí stránce je dále dostupný denní graf průjezdů vozidel z předchozího dne a snímek z radarové stanice v rozlišení 320x180 pixelů, který je, za předpokladu správného fungování radarové stanice, aktualizován každých několik minut. U některých měst jsou pak v dolní části zobrazeny aktuální informace z tamní meteorostanice [32]. V tomto bloku zmíněné statistické výstupy by mohly v některých případech rovněž posloužit jako zdroj dat pro simulaci silniční dopravy, ale hlouběji se tím práce nezabývá.

Dalšími statistikami, které tento dopravní portál nabízí, jsou grafy jednotlivých okresů, jmenovitě se jedná o okresy Domažlice, Klatovy a Tachov. Jsou to souhrnné informace průjezdů vozidel po dnech, stejně jako o tytéž informace rozdělené podle typů vozidel, které se v tomto případě dělí do čtyř kategorií: osobní vozidla, dodávky, nákladní vozidla a kamionové soupravy. Dále je zde uvažována průměrná rychlost po dnech. Ve zbylých záložkách jsou ukázky účinných zařízení, s vysokým procentem zpomalení, a celkový přehled intenzity všech zařízení rozdělených po dnech [35].

Lze zde také nalézt informační mapu dopravních omezení a nehod, která je založena na otevřené mapě OSM, stejně jako jsou po kliknutí na konkrétní značku zobrazeny stručné informace, o jakou komplikaci se jedná.

Na stránce aktualit je pak průběžně aktualizovaný seznam nově zapojených obcí, ale také seznam zařízení, které jsou po nějaký čas mimo provoz, doplněné o časovém rozsahu této komplikace.

Posledním a stěžejním bodem jsou vlastní otevřená data (open data) nabízená v rámci portálu v „*maximálním možném rozsahu a podobě, v jaké byla vytvořena a v jaké jsou spravována*“ [36]. Data jsou nabízena k volnému použití za předpokladu zachování odkazu na původní stránku a jsou uvolňována pod Open Database Licence. Jedná se o archivy komprimované do formátu ZIP, které jsou ke stažení dostupné v několika podobách. Je zde možnost stáhnout souhrnná data pro celý rok 2015 a dále možnost stáhnout data po týdnech, nebo dnech pro data od začátku roku 2016 až do současnosti s patřičným časovým rozdělením, kdy jsou tato data průběžně aktualizována. Jedná se o data, která jsou v souladu se současnou legislativou patřičně anonymizovaná pro použití širokou veřejností [32]. Těmto datům a formátům a obsahu jednotlivých souborů se podrobněji věnuje Kap. 6.

5.2.2 Ostatní podobné portály

V rámci projektů s podobným zaměřením, tedy pomocí monitorování dopravy motivovat řidiče k ukázněné jízdě, zprostředkovaných společností GEMOS CZ lze nalézt další potenciální zdroje informací.

Jedním z větších webových portálů je [30]. Jedná se o informace přímo od společnosti GEMOS. Struktura a členění této stránky je téměř shodná s dopravním portálem plzeňského kraje. Rozdíl je však obsahový. Tato stránka nabízí k jednotlivým zapojeným městům pouze data ve formátu PDF ve formě měsíčních statistik, nebo ve formě grafového výstupu po dnech. Otevřená data k těmto místům nejsou dostupná a v současné době se připravují. Tato data jsou také v některých případech dostupná přímo z webových stránek jednotlivých zapojených měst. Z tohoto portálu lze navštívit v rámci mapy radarových stanic také kamery, které se nacházejí podél českých dálnic. Tyto však nabízejí pouze statická obrazová data pořízená přímo ze stanice, kdy je náhledový obrázek aktualizován spíše nahodile s periodou několika, mnohdy i desítek, minut [30].

Dalším portálem na stejném základu je dopravní portál pro Středočeský kraj. Ten nabízí, obdobně jako plzeňský portál, měsíční a případně denní statistiky. Dále je možné navštívit stránku s otevřenými daty, která jsou zde členěna rovněž buďto po roce, týdnu, nebo dnu [31].

Tento portál zprvu neposkytoval žádné archivy, ale pouze stručné statistiky k jednotlivým záznamovým bodům v podobě obrázků přímo v náhledu jednotlivých radarových stanic. V průběhu roku 2018 však byla postupně doplněna data od 3. 9. 2017 a od té doby jsou nahrávána pravidelně. Archivy z tohoto zdroje se dále liší strukturou. Ačkoliv obsahují stejný informační soubor `readME`, jako ten na plzeňském portálu, dokonce na konci odkazující na stránky o dopravě plzeňského kraje, obsahují kromě CSV souborů s lokalitami stanic a vlastních záznamů také soubor, který poskytuje informace o vahách, které jsou k některým tamním radarovým stanicím přiřazeny [31]. Tento soubor s největší pravděpodobností poskytuje informace o dalším cizím klíči, které tyto váhy používají, avšak po zběžném průzkumu některých archivů se nepodařilo určit jejich účel, ani případná další data, která by byla díky těmto vahám poskytována. Tímto zdrojem se práce dále hlouběji nezabývá.

6 Data z dopravního portálu plzeňského kraje

Tato část práce se zaměřuje na formát, v jakém jsou data z kamerových a radarových stanic uložena online. Dále také na to, jakým způsobem se liší různé typy dostupných archivů, zda a jak se liší jejich struktura, případně struktura jednotlivých záznamů. Nakonec obecně popisuje obsah jednotlivých souborů v archivech.

6.1 Formát uložení

Jak již bylo zmíněno v předchozí kapitole této práce, data jsou dostupná ve třech typech archivů, které jsou ukládány ve formátu ZIP, což je rozšířený a obecně užívaný formát pro kompresi souborů. Na webových stránkách dopravy plzeňského kraje jsou k archivům dostupné následující informace: *název souboru*, *typ souboru*, *velikost souboru* a *datum poslední změny*. Podle těchto parametrů je také možné tyto archivy řadit. Každý typ archivu má vyhrazen vlastní adresář, ze kterého je možné soubory stahovat. Tyto archivy jsou rozděleny podle periody, za kterou byla data sbírána. První možností, s nejkratší periodou sběru, je *jeden den*, druhou možností jsou data shromážděná za *jeden týden*. Třetí a poslední možností pak data sebraná za *jeden rok*.

V případě archivu obsahujícího informace za jeden den, je název souboru v následujícím formátu:

DOPR_D_rrrrmdd.zip

Úvodní část DOPR_D_ je pro všechny archivy tohoto typu totožná a písmeno D reprezentuje označení denního sběru. Následuje unikátní označení dne, kdy byla dopravní data nasbírána, ve formátu *rok*, *měsíc* a *den* bez mezer, přičemž čísla reprezentující den a měsíc jsou v každém případě dvouciferná a jsou případně doplněna o nulu na začátku (např. leden je označen 01). Jméno archivu je pak standardně zakončeno koncovkou ZIP souboru.

V případě, že se jedná o archiv obsahující informace za jeden týden, je název souboru v následujícím formátu:

DOPR_T_rrrrt.zip

Jako u předchozího formátu je první část jména vždy stejná DOPR_T_, kde písmeno T reprezentuje označení týdenního sběru. Následuje unikátní časové označení daného archivu. To je složeno z označení kalendářního *roku*, které je doplněné o číslo *týdne*

v roce. V tomto případě však není případná jednociferná hodnota doplněna na počátku nulou, v takovém případě je tedy název souboru o znak kratší. První týden roku 2017 byl označen `DOPR_T_20171.zip`, padesátý druhý týden pak `DOPR_T_201752.zip`. Název je opět zakončen standardní koncovkou pro ZIP soubor.

Ačkoliv je v případě ročního sběru dat online dostupný pouze jeden archiv, lze předpokládat, že se jedná o následující formát názvu:

`DOPR_R_rrrr.zip`

Obdobně je zde neměnná úvodní část názvu `DOPR_R_`, kde písmeno R reprezentuje označení ročního sběru. Následuje označení kalendářního roku a standardní koncovka ZIP souboru.

6.2 Struktura a obsah archivů

Po otevření jednotlivých typů archivů lze zjistit, že všechny obsahují stejnou trojici souborů.

V prvé řadě se jedná o soubor `readME`, který je v každém typu archivu ve formátu TXT, což je formát prostého textu, a se stejným obsahem. Tento soubor je možné nalézt na přiloženém CD v adresáři *A – Testovací vstupní data*.

Tento soubor poskytuje stručné informace o projektu, v rámci kterého jsou otevřená data poskytována a zároveň informace o typu zařízení, které záznamy pořizuje. Sběr dat je součástí česko-bavorského projektu *Klidné příhraničí* a data jsou pořizována zařízeními *SYDO Traffic Zeus*. Podrobnější informace o záznamových stanicích jsou dostupné na stránkách výrobce⁴ a v případných informačních letácích⁵.

Dále také informuje o licencích, pod kterou jsou otevřená data poskytována. Ta jsou dostupná pod Open Database License⁶ a Database Contents License⁷.

Následují stručné vysvětlivky k souborům s daty. Nejdříve je zmíněn formát názvu obou souborů. V případě souboru s vlastními záznamy se jedná o stejnou strukturu jména, jako u archivů popsanych v Kap. 6.1, s tím rozdílem, že zde se nejedná o soubor

⁴ <http://www.gemos.cz/>

⁵ <http://www.intelligentniukazatel.cz/data/letak-mereni-rychlosti.pdf>

⁶ <https://opendatacommons.org/licenses/odbl/1.0/>

⁷ <http://opendatacommons.org/licenses/dbcl/1.0/>

ve formátu ZIP, ale CSV. Druhý soubor, s výčtem všech záznamových stanic a informacemi o nich, je v souboru `readME` označen jako `Location.csv`, přičemž by toto jméno mělo pro všechny archivy stejné. Po průzkumu archivů je možné zjistit, že skutečně užívané jméno tohoto souboru je `Locations.csv`. Toto jméno je pak pro všechny archivy opravdu neměnné.

Nakonec jsou popsány názvy atributových sloupců obou souborů (podrobněji v následující podkapitole), definovány primární klíče obou tabulek, a je vysvětleno, jakým způsobem je vytvořen cizí klíč mezi oběma pomyslnými tabulkami, což je doplněno o několik vysvětlivek k jiným sloupcům.

Vzhledem k totožnému obsahu archivů a dále jednotnému formátu vlastních dat napříč všemi typy archivů bude v následujících bodech kapitoly hovořeno o souborech a archivech souhrnně, přičemž uvedené příklady jsou ze souborů, které uchovávají data s denní periodou s tím, že popsané postupy je možné aplikovat stejným způsobem i na zbylé typy archivů.

6.3 Obsah a význam CSV souborů

V této kapitole je stručně vysvětlen formát CSV a význam jednotlivých sloupců z obou souborů. Jako primární zdroje informací slouží soubor `readME.txt` spolu s informacemi poskytnutými na stránkách plzeňského kraje, které jsou doplněny informacemi o typologii vozidel získanými od kontaktní osoby pro otevřená data v podobě stručného seznamu (tento seznam odpovídá Tab. 4). Správnost těchto informací byla ověřena porovnáním poskytnutých informací vůči předpokládaným hodnotám. V případě chybějících informací o bližším účelu atributu jsou popsány jeho obvyklé hodnoty a proveden odhad možného významu na základě kontextu vůči hodnotám z ostatních sloupců.

6.3.1 Formát CSV

Formát *Comma-Separated Values* (hodnoty oddělené čárkami), zkráceně CSV, je rozšířený a jednoduchý textový formát určený zejména pro výměnu tabulkových dat.

Základem tohoto formátu jsou atributové sloupce, jejichž počet se v rámci souboru nemění, které jsou od sebe odděleny jednotným oddělovacím znakem, nejčastěji, jak již název napovídá, čárkou. Jednotlivé záznamy jsou pak od sebe odděleny znakem, nebo značkou konce řádku. Každý CSV soubor také může obsahovat hlavičkovou řádku,

kteřá určuje pojmenování atributů. Důležitý je také výskyt uvozovek, které slouží jako takzvaný *escapovací* znak. Je možné jimi ohraničit atribut v případě, že je třeba v jeho obsahu použít některý z řídicích znaků, což je v tomto případě oddělovací znak, případně konec řádky. Pokud by text měl obsahovat uvozovky, jsou zdvojeny [33].

Soubory z dopravního portálu plzeňského kraje se od standardu CSV liší hned v několika bodech. Hlavičkové řádky nejsou uvedeny v rámci souborů, nýbrž v rámci popisného `readME` souboru a jsou doplněny o několik dalších informací. Jako standardní oddělovač zde není využita čárka, ale svislá čára, tedy `|`. Vlastní atributy, zejména pak textové řetězce, jsou ohraničeny ASCII verzí uvozovek, tedy `"`. V rámci jedné z hodnot, která obsahuje data z dvoudimenzionálního pole, je navíc využita dvojice separátorů. Nejprve jsou data rozdělena středníkem, tedy `;`, následně jsou hodnoty rozděleny čárkou, tedy `,`. Desetinná čísla využívají desetinnou tečku.

6.3.2 Locations

Soubor `Locations.csv` se skládá ze záznamů, které jsou složeny celkem z pěti atributů popsaných v souboru `readME` následovně:

```
Name|Town|Street|IdDevice|IdArea
```

Primárním klíčem této tabulky je zvýrazněná hodnota `IdDevice`.

Po nahlédnutí do souboru lze zjistit, že hodnoty všech pěti sloupců jsou navíc ohraničeny uvozovkami.

Jeden z možných záznamů pak vypadá například takto:

```
"Česká Kubice, směr od Německa"|"Česká Kubice"|"Česká Kubice"|  
"KP055"|"5"
```

Následuje popis jednotlivých atributů.

IdDevice

Jedná se o primární klíč této tabulky ve formátu:

```
KPxxx
```

Dvojice znaků `KP` je vždy stejná a lze předpokládat, že je zkratkou pro označení kraje, v tomto případě tedy Plzeňský kraj, neboť v případě otevřených dat pro Středočeský kraj je tato zkratka `SK`. Dále následuje trojice dekadických čísel, která slouží k jednoznačné identifikaci záznamového zařízení v rámci otevřených dat.

Číslování těchto hodnot začíná od čísla 51, přičemž další hodnoty se zvyšují, ale ne vždy na sebe přímo navazují.

Name

Tento atribut je slovním označením konkrétní radarové stanice a hodnota tohoto sloupce je dále využívána na webových stránkách dopravy plzeňského kraje pro označení radarové stanice na interaktivní mapě.

Town

Jméno obce, ve které se daná radarová stanice nachází.

Street

Jméno ulice, ve které se daná radarová stanice nachází. V případě, že se jedná o malé obce, které ulice nemají, je zde opět použito jméno obce. Například hodnota Česká Kubice u obce Česká Kubice a hodnota Plzeňská u obce Tachov

IdArea

Jedná se o číselné označení oblasti, kde se daná radarová stanice nachází. Tato hodnota je shodná pro všechny záznamové stanice, které se nacházejí v jedné obci.

6.3.3 Dopravní data

Jméno tohoto souboru je ve formátu `DOPR_D_rrrrmdd` a je kromě koncovky shodné se jménem archivu, ze kterého pochází. Obsahuje celkem jedenáct atributů, které jsou vyjmenovány v souboru `readME`, některé i stručně vysvětleny. Atribut `RychlostHistorie` se pak v některých případech skládá z pole hodnot proměnlivé délky. Celková podoba jednoho záznamu je pak následující:

```
IdDetektor|DatumCas|Intenzita|IntenzitaN|Obsazenost|  
Rychlost|Stav|TypVozidla|Trvani100|RychlostHistorie|  
TypVozidla10
```

Jako primární klíč zde slouží dvojice zvýrazněných hodnot `IdDetektor` a `DatumCas`.

V průběhu času se vlastní formát jednotlivých záznamů nemění, mění se však počet stanic, které záznamy pořizují, a způsob, jakým jsou tyto záznamy seřazeny. V některých případech jsou data řazena podle času, kdy byly jednotlivé záznamy

pořízeny, jindy jsou nejdříve seřazeny podle toho, z jaké stanice záznam pochází a až poté podle času pořízení.

Následuje popis jednotlivých atributů.

IdDetektor

Jedná se o osm číslic v následujícím formátu:

10XXXXYYZ

První dvě čísla jsou v tomto případě neměnná konstanta 10. Následující trojčíslí je pak číselná část hodnoty *IdDevice* ze souboru *Locations*, která jednoznačně identifikuje radarovou stanici v systému a slouží tak jako cizí klíč. Další dvojčíslí, podrobněji popsané v Tab. 1, identifikuje, jakým způsobem byl záznam pořízen.

Tab. 1: Označení typu detekčního zařízení

00	radarový záznam
10	video detekce
20	kamerový záznam

Tato typologie záznamových zařízení je stručně zmíněna na stránkách výrobce zařízení a v příslušných brožurách, ale tento popis neobsahuje informace o účelu daných zařízení, ani informace o technické specifikaci. Rozdíly mezi těmito záznamy jsou hlouběji analyzovány v dalších bodech práce.

Poslední číslo stanovuje, jakým směrem vůči zařízení se vozidlo pohybovalo. Označení směrů je popsáno v Tab. 2.

Tab. 2: Označení směru pohybu vozidla

1	ve směru
2	v protisměru

DatumCas

Jedná se o textový řetězec, který je stylizován do následujícího formátu:

RRRR-MM-DD hh:mm:ss.sss

První částí je identifikace dne v roce ve formátu rok-měsíc-den, přičemž hodnoty jsou odděleny pomlčkami. Následuje mezera a identifikace času v rámci dne s předností na

tisíciny vteřiny. Hodnoty jsou od sebe odděleny dvojtečkou, milisekundy pak desetinnou tečkou.

Intenzita

Jedná se o neměnnou hodnotu, která je vždy 1, a jejíž účel není v dostupných zdrojích popsán a není ani zřejmý.

IntenzitaN

Jedná se o pole, které je využito v případě, kdy je jako způsob záznamu použita video detekce. V takovém případě tato hodnota s největší pravděpodobností napomáhá při určení typu vozidla. Kupříkladu hodnota 3.0 odpovídá typu vozidla 8, 2.0 typu 6, 1.5 typu 4 a hodnota 1.0 pak typu 2.

U ostatních forem detekce se jedná vždy o hodnotu 1.0.

Obsazenost

Jedná se o neměnnou hodnotu, která je u všech záznamů 100.00.

Rychlost

Jedná se o číselnou hodnotu s přesností na dvě desetinná místa, která jsou oddělena desetinnou tečkou. K označení hodnoty rychlosti je však vždy použita pouze celočíselná část.

V případě, že se podařilo zaznamenat průjezd vozidla, ale nebylo možné určit rychlost, vyskytuje se zde v záznamu nulová hodnota, tedy 0.00.

Stav

Ve všech případech se jedná o neměnnou hodnotu 0.

TypVozidla

Číselná hodnota nabývající čtyř standardních hodnot na základě toho, o jaký typ vozidla se jedná. Popsáno v Tab. 3.

Tab. 3: Jednodušší typologie vozidel

1	motocykl
2	osobní automobil
3	dodávka
4	nákladní vůz

Trvani100

Jedná se o hodnotu, která je využívána pouze v případě, že byl záznam pořízen video detekcí. Pro zbylé dva typy záznamů je vždy nulový. Význam tohoto pole není nikde vysvětlen, ale vzhledem ke kontextu se s největší pravděpodobností jedná o dobu, po kterou byl záznam pořizován, kde číslo 100 v názvu značí stovky milisekund. Číslo 5 by tedy znamenalo 500ms.

Ačkoliv by se mohlo zdát, že tento záznam koresponduje s atributem *RychlostHistorie*, není tomu tak. V případě šestého záznamu ze souboru *DOPR_D_20160103* je tato hodnota 5, avšak záznamů je celkem 10, u devátého je pak tato hodnota 20 a záznamů je 6.

RychlostHistorie

Tento záznam v podobě textového řetězce je opět využit pouze v případě, kdy se jedná o video detekci. V takovém případě se jedná o dvourozměrné pole proměnlivé délky, které v první hodnotě obsahuje index pole začínající nulou a v druhé hodnotě obsahuje hodnoty naměřené rychlosti. Tyto hodnoty jsou vzájemně odděleny čárkou, dvojice jsou pak odděleny středníky, přičemž se středník zároveň vyskytuje i za poslední hodnotou v záznamu.

Vztah mezi tímto atributem a atributem *Rychlost* je, že se hodnota rychlosti rovná hodnotě rychlosti s indexem 0 v atributu *RychlostHistorie*.

Ukázkový záznam po odstranění uvozovek:

```
0, 43; 1, 48; 2, 50; 3, 52; 4, 53; 5, 55; 6, 56; 7, 56; 8, 52; 9, 53; 10, 53;
```

TypVozidla10

Podobně jako u hodnoty *TypVozidla* se i zde jedná o určení typu vozidla na základě pořízeného záznamu. Tato hodnota však poskytuje přesnější rozdělení, které je popsáno v Tab. 4.

Tab. 4: Podrobnější typologie vozidel

1	motocykl
2	osobní automobil
3	osobní automobil s přívěsem
4	dodávka
5	dodávka s přívěsem
6	lehký nákladní vůz
7	lehký nákladní vůz s přívěsem
8	kamion
9	kamion s přívěsem
10	autobus

7 Transformace dat

Tato kapitola se zabývá způsoby, jak z dříve zmíněných archivů získat data relevantní pro simulaci silniční dopravy a jak tato data transformovat do vhodnějšího formátu odstraněním redundantních a nepotřebných informací. Dále navrhuje několik možných přístupů k transformacím s ohledem na přístup k datové agregaci a účelu výsledných dat.

7.1 Výsledný formát

Jako výstupní formát transformací je rovněž zvolen formát CSV, zejména díky jeho jednoduché struktuře a nízké náročnosti pro případné další zpracování databázovými nástroji, nebo obdobným softwarem.

7.1.1 Pravidla výsledného CSV formátu

Tento nový formát rovněž neobsahuje hlavičkovou řádku, ale případné informace o konkrétním způsobu transformace poskytuje buďto ve svém názvu, nebo v příložených metadatech. Formát názvu tohoto souboru je následující:

```
RRRRMMDD - N - info.csv
```

První částí je identifikace dne, hodnota N pak odpovídá typu transformace a `info` obsahuje stručné, bližší informace o specifikacích dané transformace.

Pro oddělení jednotlivých datových atributů je použit středník. V případě záznamu desetinného čísla je použita desetinná tečka. Je-li potřeba využít v rámci záznamu speciální znak, pak je tento atribut uvozen ASCII verzí uvozovek. Je-li třeba užít uvozovky, jsou pak zdvojeny.

Při užití agregace dat se pak jako atribut může vyskytovat pole více hodnot. Toto pole je uvozeno hranatými závorkami a uvnitř jsou jednotlivé hodnoty odděleny opět středníkem.

7.2 Očištění dat

Jak již bylo zmíněno v předchozích kapitolách, aby bylo možné využívat volně dostupná data pro simulaci silniční dopravy, je třeba provést jejich vhodnou

transformaci, jejímž účelem je výběr informací, jako je hustota dopravy, rychlost vozidel a typologie vozidel nebo řidičů.

V případě těchto otevřených dat je prvním krokem určení, se kterými poskytnutými atributy je vhodné dále pracovat vzhledem k možnosti jejich využití v simulaci silniční dopravy. Tato data lze pro přehlednost rozdělit do několika definujících skupin, které jsou popsány v následujících podkapitolách.

7.2.1 Způsob pořízení záznamu

Podle části atributu `IdDetektor` lze jednotlivé záznamy z původních dat rozdělit do tří kategorií na základě způsobu, jakým byla pořízena. Jedná se o radarový záznam, video záznam a kamerový záznam.

Tyto typy záznamů se liší tím, jaké informace jsou schopné poskytovat a s jakou přesností. Všechny typy záznamů jsou schopné poskytnout informaci o přesném čase, ve kterém byly pořízeny.

V případě záznamů pořízených pomocí radaru se dále jedná o spolehlivou informaci o rychlosti projíždějícího vozidla v obou směrech měření.

U záznamů pořízených video detekcí se v případě záznamu ve směru k měřicí stanici jedná o záznamy s historií rychlosti vozidla a také s hodnotou, jak dlouho byl záznam pořizován v řádu stovek milisekund. V případě pohybu od zařízení je pořízena pouze informace o tom, jak dlouho video záznam probíhal. V obou případech pak také dochází k určení typů vozidel. K relevantnímu určení rychlosti vozidla je v tomto případě nutné, aby daná stanice obsahovala také funkční radar. U stanic, které radarový záznam nepořizují, je možné si všimnout velkého množství nulových záznamů rychlosti i u video detekce.

Posledním typem záznamu je kamera. Tento typ záznamu pořizuje informace pouze o průjezdech ve směru ke stanici. V tomto případě se může jednat o radarový záznam, který byl zároveň také doplněn fotografií vozidla. Důvod k jejímu pořízení mohlo být překročení maximální rychlosti nastavené v konkrétní stanici. Avšak, jedná se pouze o domněnku. S jistotou lze říci, že tento způsob pořízení záznamu je schopen určit rychlost projíždějícího vozidla a to i u stanic, kde nejsou pořizovány radarové záznamy. V některých případech je použit i při určení rychlosti vozidla v případě video záznamu.

Možným problémem ohledně způsobu pořízení záznamu je možnost vícenásobného zaznamenání jednoho vozidla více způsoby. Při průzkumu archivů bylo, zejména v hodinách s nízkou hustotou dopravy, možné nalézt několik podezřelých záznamů, ale zároveň také mnoho případů, kde byl pořízen pouze jeden typ záznamu s odstupem několika desítek vteřin. Možným řešením by tak bylo záznamy ze stejného zařízení a směru v rámci krátkého časového úseku (např. 0,5 vteřiny) sjednocovat do jednoho. Touto problematikou se práce hlouběji nezabývá a v transformacích není zohledněna.

7.2.2 Umístění

Prostorové dělení dat je stěžejní pro vytvoření realistických generačních bodů, což je zajištěno identifikací místa, kde byl záznam pořízen. Tuto informaci je u otevřených dat možné zjistit pomocí dělení záznamových stanic v souboru `Locations` a to jak jejich jednoznačným určením pomocí identifikačního čísla, tak případným označením širší oblasti, kde se záznamová stanice nachází. Přesné informace o poloze dané stanice, například GPS souřadnice, je pak možné získat z mapy, která se nachází na webových stránkách dopravy plzeňského kraje. Název této stanice pak koresponduje s názvem v souboru `Locations`.

Dalším způsobem prostorového dělení dat je v souboru se záznamy průjezdů, kde lze využít část atributu `IdDetektor` k určení směru pohybu vozidla. Z tohoto atributu je navíc možné získat také informace o konkrétním typu měřícího zařízení, které záznam pořídilo.

7.2.3 Čas

Důležitou informací je rovněž možnost jednoznačně časově zařadit pořízený záznam. To je v nejhrubší podobě zajištěno již výskytem souborů se záznamy. Jednalo by se tak o dělení po dnech, týdnech, nebo letech.

Jednotlivé záznamy je pak možné do časového kontextu zařadit pomocí informací ve sloupci, který je v původních datech označen jako `DatumCas`. Z tohoto atributu je navíc možné, v závislosti na časovém rozsahu výsledných souborů, některé prvky časové identifikace vypustit. Například, pokud by se v souboru vyskytovala data pouze z jednoho dne, není potřeba pro identifikaci záznamu používat celé datum, ale stačí pouze časová identifikace v rámci dne.

7.2.4 Rychlost

Rychlost projíždějících vozidel je další užitečnou informací, zejména pak u mikroskopických, nebo nanoskopických pojetí simulace silniční dopravy, kde napomáhá při vytvoření typologie vozidel a řidičů.

V případě těchto dat se však vyskytuje několik problémů, které je třeba mít při dalším zpracování na paměti, neboť jejich opomenutí poskytne nerealistický pohled na věc.

Jedním z nich je účel a místo pořízení těchto dat. Záznamové stanice se nacházejí na hranicích obcí a účelem projektu, v rámci kterého byly instalovány, je zajištění dodržování maximální povolené rychlosti ve městech. Stanice jsou navíc dobře viditelné a jasně označené, což může v některých případech vést ke zkreslení rychlosti, kterou by se někteří řidiči zbytkem daného úseku reálně pohybovali.

Možným řešením je data o rychlosti vypustit a pro určení rychlosti vozidel využít informace o maximální povolené rychlosti na dané komunikaci s tím, že v rámci zachování plynulosti provozu se řidiči snaží držet co možná nejbližší této hodnotě. V takovém případě však dojde ke ztrátě informací o případném rozdílném chování řidičů v bezprostřední blízkosti zařízení, což je pro simulaci rovněž důležité, a zároveň to nemusí být způsobeno jen přítomností měřicí stanice, ale také například i sníženou viditelností, nebo zhoršeným stavem vozovky.

Další informací, kterou je třeba zohlednit, je atribut `RychlostHistorie` pořizovaný zařízeními s video detekcí, kdy dochází k několikanásobnému zaznamenání rychlosti projíždějícího vozidla.

Zde se však setkáváme s problémem, kdy jsou tato data dostupná pouze u záznamů pořizovaných video detekcí a zároveň pouze u průjezdů ve směru k měřicí stanici, neboť v opačném směru nejsou tyto informace pořizovány. Z tohoto důvodu není tento atribut dále využíván a pro určení rychlosti je použit standardně atribut `Rychlost`.

7.2.5 Typologie vozidel

Poslední užitečnou skupinou dat, využitelnou v simulaci, je určení typu vozidla a to opět zejména v podrobnějších přístupech, jakými je mikroskopické a nanoskopické pojetí simulace.

Otevřená data nabízejí dva způsoby rozdělení. Jedním je zjednodušené rozdělení do čtyř kategorií, druhým je rozdělení do celkem deseti kategorií. Zde je však nutné přihlídnout k tomu, že v samotných datech jsou v podrobnějším rozdělení reálně využity pouze čtyři kategorie, které korespondují s kategoriemi užitými v PDF souborech na stránkách dopravy, a v případě jednoduššího členění tři. Ačkoliv je kategorie *motocyklů* přítomna v obou typech rozdělení, není reálně využita. Dále se dvě z kategorií překrývají. Jedná se o typy *osobní automobil* a *dodávka*. Poslední zjednodušená kategorie *nákladních vozů* je pak v podrobnějším pojetí rozdělena na *lehké nákladní vozy* a *kamiony*.

7.3 Agregace dat

Stěžejním bodem agregace pro využití dat v oblasti simulace silniční dopravy je rozdělení do časových bloků vhodné velikosti, neboť užití dat v podobě jednotlivých záznamů není jako vstupu pro generační funkci, bez použití další obsáhlé transformace, vhodné a možné. Zároveň by takový způsob zpracování znamenal přesnou replikaci průjezdů v dané oblasti, což pro simulaci není v naprosté většině případů potřeba.

K agregaci však lze přistupovat i z jiného hlediska, než jen časového. V případě těchto dat je dalším možným způsobem agregace případné sjednocení obou směrů průjezdů kolem stanice, nebo sjednocení dopravních dat z více zařízení podle oblasti, respektive obce, ve které se nacházejí.

Posledním možným způsobem agregace je v případě těchto dat sjednocení všech typů vozidel, neboť některé přístupy k simulaci silniční dopravy tuto informaci nevyužívají. Zároveň se jedná o nejsnazší řešení problému, kdy záznamová zařízení určují typ vozidel pouze ve chvíli, kdy se jedná o pořízení záznamu video detekcí.

7.3.1 Bez použití agregace

Zpracování bez použití agregace je dobrým způsobem pro zeštíhlení otevřených dat o informace, které nemají pro simulaci silniční dopravy hodnotu, a zlepšení jejich přehlednosti. Dalším důležitým faktorem tohoto přístupu je eliminace chybných záznamů, případně eliminace záznamů, které neposkytují užitečná data pro konkrétní typ transformace. V případě těchto dat se může jednat například o eliminaci záznamů pořízených typem zařízení, které není schopné spolehlivě určit rychlost, nebo typ vozidla.

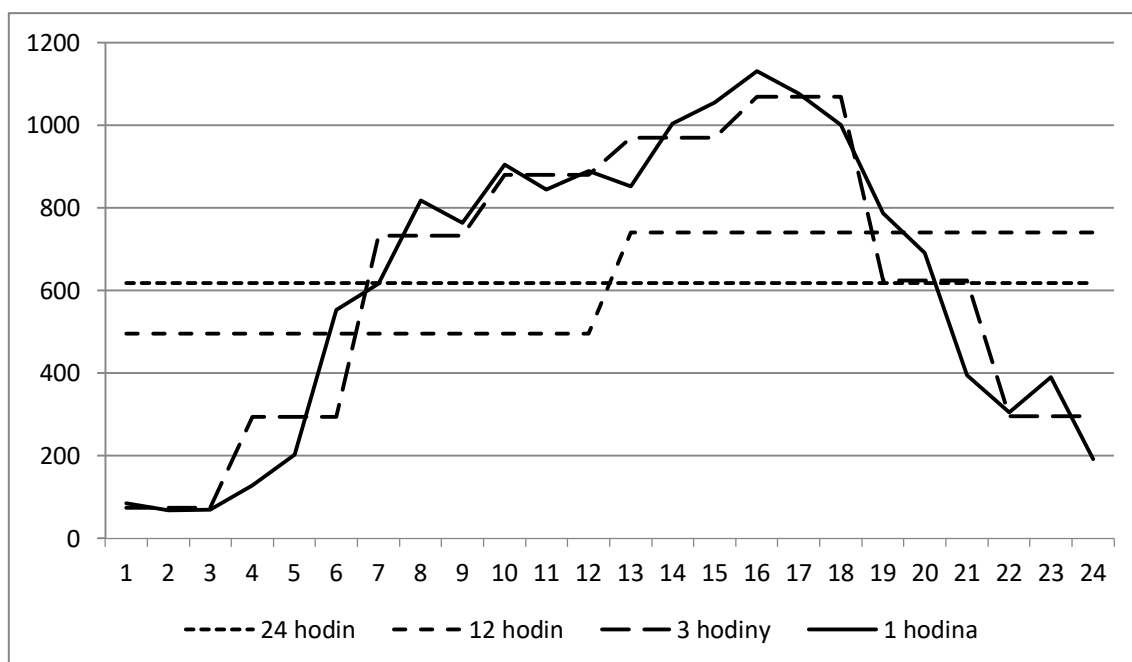
Využití této metody je však pro využití v simulaci spíše vhodným mezikrokem, jehož účelem je pročištění dat a je vhodné tento přístup kombinovat s některými dále zmíněnými agregačními metodami, nebo jinými.

7.3.2 Agregace podle času

V simulaci silniční dopravy se pro generování nových vozidel, nebo objektů, které jim odpovídají, využívá různých metod, pro které může být důležitou informací předpokládaný počet průjezdů za určitý časový úsek. Tato hodnota je dále randomizována, aby co nejvíce napodobila podobu dopravy popsanou počty průjezdů, ale zároveň zohlednila nahodilost a nepředvídatelnost konkrétních průjezdů.

Agregování dat do časových bloků je v tomto ohledu stěžejním postupem pro další využitelnost dat v oblasti simulace silniční dopravy.

Další otázkou k řešení je velikost užitých časových bloků. Využití příliš malých bloků může působit nepřehledně, v případě řídké dopravy v některých případech paradoxně ještě zvětšit vstupní data. Další nevýhodou malých časových úseků je malý stupeň volnosti při vytvoření nahodilosti v datech, což může vést k vzniku obdobného problému, jako v případě, kdy by agregace použita nebyla. Naopak, v případě použití příliš velkých časových kroků není možné danou dopravní situaci popsat dostatečně přesně a dojde ke ztrátě stěžejních informací, jako například lišící se hustota a složení dopravy v průběhu různých dnů. Ztrátu této přesnosti popisuje Obr. 1, jsou použita data ze stanice č. 51 na průjezdech ve směru ze dne 2. 11. 2018.



Obr. 1: Vliv velikosti časových bloků na zkršení podoby dopravy

Velikost časového bloku se v tomto případě může pohybovat v rozsahu několika desítek minut, do několika hodin v závislosti na použitém typu simulace, nebo simulačním softwaru. V případě, že jsou data o podobě dopravy dostupná z jiného zdroje, nebo jsou generována jiným způsobem, je možné data o průjezdech agregovat do bloků velikosti simulovaného období (např. dne).

Při určení časového bloku vhodné velikosti je dále možné, průzkumem dat z velkého množství dnů, vytvořit jakousi typologii dnů, která může být dále využita při transformaci dat počtů průjezdů za delší časová období z jiných, méně přesných zdrojů pro získání přesnější podoby dopravy.

7.3.3 Agregace podle místa

Využití podrobnosti mnoha potenciačních generačních míst se dvěma směry sběru je vhodné zejména v případě simulace malých oblastí, například obce a přilehlého okolí. Při simulaci větších oblastí je možným řešením agregování dat průjezdů v rámci obce do jedné hodnoty, nebo zastřešující kategorie a to zejména u menších obcí, u kterých výskyt měřících zařízení na okrajích významných komunikací poskytuje komplexní a realistický obraz celkové hustoty dopravy v jejím rámci.

Důležitým faktorem, který je třeba zohlednit v tomto případě i v případě, kdy tato agregace neprobíhá, je zejména v případě menších obcí relativní blízkost měřících

zařízení, což může v kombinaci s malým množstvím alternativních cest pro vozidla znamenat umělé nabobtnání dopravy v oblasti, neboť dochází k dvojitmu započtení daného vozidla. Nejprve je vozidlo zaznamenáno při příjezdu do obce ve směru k zařízení a poté je zaznamenáno při odjezdu z obce ve směru od zařízení jinou stanicí.

Možným řešením je nevyužívat data z protisměru. Tento přístup navíc podporuje fakt, že pořízení záznamu pomocí kamery je schopné zachytit vozidla pouze ve směru k měřicímu zařízení, což může způsobit zkreslení počtu průjezdů oproti protisměru. Pro ověření této informace a obecně schopnosti zaznamenávat průjezdy v protisměru je však třeba dalšího průzkumu.

Tento způsob agregace není v Kap. 8 samostatně demonstrován, neboť by se jednalo o totožný postup transformace, avšak informace by nebyly děleny podle jednotlivých stanic, ale kupříkladu podle identifikačního čísla oblasti, nebo vlastního způsobu dělení na základě blízkosti stanic na mapě.

7.3.4 Agregace typů vozidel

Ačkoliv může být typ vozidla u některých typů simulace užitečnou informací, pro mnoho jiných je tato informace redundantní, neboť se v daných přístupech chovají všechna vozidla stejným způsobem, nebo se s jednotlivými vozidly blíže nepočítá. V takových případech je tedy vhodným řešením zaznamenané záznamy zařadit pouze do jedné generické kategorie, která zachycuje celkový počet průjezdů.

7.4 Atributy výsledného formátu

V této podkapitole jsou obecně popsány možné výstupní formáty dat z provedených transformací s ohledem na jejich typ a fakt, že ne všechna data jsou využita ve všech typech transformací a ne vždy mají tato data stejný formát.

Obecný výsledný formát záznamu po transformaci pak má následující podobu:

```
id_zařízení;id_čas{;rychlost}{;počet_vozidel}{;typ_vozidla}
```

Konkrétní podoba jednotlivých atributů je popsána v následujících bodech. Atributy ve složených závorkách jsou nepovinné parametry, které nemusí být v některých případech využity.

7.4.1 Identifikace místa

Původní označení místa, kde byl záznam pořízen, je v souboru s daty v následujícím, již dříve zmíněném, formátu:

```
10XXXYYZ
```

Z těchto dat je pro jednoznačné určení místa záznamu nutné využít pouze trojčíslí označené znaky XXX. K podrobnějšímu rozdělení lze využít poslední číslo, označené Z, které určuje směr, kterým se vozidlo pohybovalo vůči záznamovému zařízení.

Identifikace místa tak má v novém formátu následující podobu:

```
id_zařízení;směr
```

V případě `id_zařízení` se jedná o trojčíselné číslo, které jednoznačně určuje záznamové zařízení. Toto číslo je shodné s označením zařízení ze souboru `Locations`. U nepovinného parametru `směr` se jedná o datový typ `boolean`, který je reprezentován v případě pohybu směrem k zařízení hodnotou 1, v opačném případě, kdy se vozidlo pohybovalo směrem od zařízení, hodnotou 0.

Dvojčíslí `YY` určující typ pořízení záznamu může být v některých případech zanedbáno. Jedná se o případy, kdy není účelem transformace data tímto způsobem členit, nebo byla data transformována tak, že došlo k eliminaci nepřesností jednotlivých typů záznamů a toto členění by tak bylo redundantní.

Pokud je tento způsob rozdělení použit, lze pro identifikaci typů záznamu využít desítkovou část původního čísla. Radarový záznam využívá hodnotu 0, video záznam hodnotu 1 a kamerový záznam hodnotu 2.

Formát pro identifikaci místa pak bude v následujícím formátu:

```
id_zařízení;typ_záznamu;směr
```

7.4.2 Časové zařazení

Určení, kdy byl záznam pořízen, je v novém formátu možno chápat třemi způsoby. V případě, že jsou záznamy stále členěny jednotlivě, a zároveň jsou ve výsledném souboru data z více dnů, jedná se o následující formát:

```
DD-MM-RRRR hh:mm:ss.sss
```

Pokud by se ve výsledném souboru vyskytovala pouze data za jeden den, jedná se o druhý, zjednodušený formát:

hh:mm:ss.sss

V obou předchozích případech jsou záznamy rozděleny s přesností na tisícinu vteřiny stejně jako v původním formátu.

Třetí a zároveň nejčastěji používaný formát, je použit v případě, že jsou jednotlivé průjezdy agregovány do záznamů podle časových bloků. V takovém případě je použita celočíselná rostoucí posloupnost začínající nulou, kde tato hodnota reprezentuje označení časového bloku. Velikost tohoto časového bloku je pak definována v rámci metadat, nebo názvu souboru.

Například při agregaci dat po třech hodinách by časový blok označený 0 reprezentoval čas od 0:00:00 do 2:59:59.999, blok označený 7 pak čas od 21:00:00 do 23:59:59.999, apod.

7.4.3 Počet vozidel

Tento atribut je využit v případě, že jsou jednotlivé záznamy nějakým způsobem agregovány. V opačném případě, kdy se jedná o jednotlivá vozidla, se tento atribut vyskytuje implicitně a jeden záznam zároveň znamená jeden průjezd.

Pokud dále nedochází k rozdělení vozidel podle typů, jedná se o jedno číslo, které určuje, kolik vozidel daným místem projelo během sledovaného úseku. Výsledný formát:

počet_průjezdů

V případě rozdělení vozidel na typy se jedná o pole hodnot, které odpovídají jednotlivým typům vozidel, buďto v absolutních číslech, nebo jako poměrový ukazatel. Formát těchto hodnot je následující:

[typ1;typ2;...;typN]

7.4.4 Rychlost vozidel

Nedochází-li k agregaci dat, určuje tento parametr rychlost zaznamenaného vozidla. Tato rychlost je uváděna v celých kilometrech za hodinu, neboť přesnější data nejsou poskytnuta.

Pokud jsou data nějakým způsobem agregována, určuje tento parametr průměrnou rychlost vozidel za sledované období. Tato rychlost je uváděna se zaokrouhlením na dvě desetinná místa s použitím desetinné tečky.

Formát tohoto atributu je následující, přičemž se jedná buď o celočíselnou hodnotu, nebo desetinné číslo:

rychlost

7.4.5 Typologie vozidel

V původních datech se tato kategorie vyskytuje hned dvakrát, avšak, jak bylo zmíněno dříve, jedná se o překrývající se informace s různou úrovní přesnosti rozdělení. Není proto třeba využívat obě tyto kategorie naráz a zároveň, vzhledem k reálnému nevyužití všech kategorií, není třeba k podrobnějšímu třídění použít všech deseti kategorií. Z těchto důvodů je v transformovaných datech využit nový způsob typologie vozidel, který kopíruje využívané kategorie přesnějšího rozdělení, a je doplněn o vlastní kategorii pro chybné, nebo nejasné záznamy. Toto rozdělení je popsáno v Tab. 5.

Tab. 5: Nová typologie vozidel

Označení	Název kategorie
1	osobní automobil
2	dodávka
3	lehký nákladní vůz
4	kamion
0	nebylo možné určit

V případě transformace využívající jednotlivých záznamů tato hodnota přímo odpovídá typu daného vozidla dle nového rozdělení.

Pokud jsou data agregována, může být tato hodnota vynechána. Je-li vynechána, pak se atribut počtu vozidel vyskytuje pouze jako jednotlivé hodnoty pro všechny typy jednotně. V případě využití kategorizace vozidel podle typů je počet vozidel vyjádřen jako pole pěti hodnot, které pořadím odpovídají číslování v Tab. 5.

Atribut s počty vozidel by pak při členění typologie vypadal následovně:

[počet_neznámých;p_osobních;p_dodávek;p_lnvožů;p_kamionů]

Velmi důležitým faktorem, ovlivňujícím tento typ informace v otevřených datech, je, že výše popsané rozdělení vozidel na jednotlivé typy se vyskytuje pouze u záznamů, které byly pořízeny video detekcí. Tento fakt demonstruje Tab. 6.

Tab. 6: Rozložení typů vozidel podle typu záznamu; stanice č. 51, ve směru

	os. auto	dodávka	leh. nákl. vůz	kamion
radarový záznam	6267	0	0	0
video záznam	3283	63	224	157
kamerový záznam	4819	0	0	0

U ostatních způsobů pořízení jsou všechna vozidla označena typem 1, tedy jako osobní automobily. Vzhledem k tomu, že původní rozdělení nemá speciální hodnotu pro neurčité, nebo chybné záznamy, lze se v tomto kontextu domnívat, že jako výchozí hodnota je právě osobní automobil.

Předchozí odstavec pak má dva jasně definované důsledky. Pro reprezentaci rozložení typů vozidel, která stanovištěm projela, je třeba stanice s video záznamem a zároveň je třeba využívat pouze tato data při práci s typy vozidel, nebo data z ostatních typů zařízení označit za chybná.

8 Demontrace popsaných transformací

V této kapitole je demonstrováno několik způsobů, jak je možné poskytnutá otevřená data transformovat pro další využití v simulaci silniční dopravy. Tyto transformace jsou dále rozděleny do několika skupin, které na sebe komplexností volně navazují.

Pro zde uvedené příklady byla využita a transformována data z 2. 11. 2018 ze souboru DOPR_D_20181102. Použitá vstupní data se nacházejí na příloženém CD v příloze A – *Testovací vstupní data*. Zároveň je přiloženo několik dalších archivů v příloze B – *Další archivy z různých časových období sběru v případě, že by tato data nebyla dále dostupná online*⁸. Ukázková data výstupů za celý den se poté nacházejí na CD v příloze C – *Ukázky transformací*.

8.1 Očištění dat

Prvním krokem a zároveň prvním způsobem transformace je očištění původních záznamů průjezdů o atributy, které nemají pro simulaci silniční dopravy žádnou informační hodnotu, a převedení relevantních atributů do vhodnější podoby.

Součástí tohoto kroku je také odstranění, nebo označení chybných, nebo nejasných záznamů. Jedná se o záznamy, které obecně neodpovídají dříve popsané struktuře, obsahují chybnou hodnotu v některém z parametrů, nebo o záznamy, které neobsahují přesné informace pro některý konkrétní typ transformace.

Popis algoritmu transformace původního záznamu do nové podoby:

1. Načtení záznamu.
2. Rozdělení původního záznamu podle dělicích znaků a eliminace nerelevantních atributů.
3. Převedení zbylých atributů do nové podoby:
 - a. rozdělení *IdDetektor* na id. číslo zařízení, informaci o směru a typu detekčního zařízení,
 - b. odstranění data z *DatumCas*, převedení do podoby, podle které je možné tyto záznamy řadit,
 - c. odstranění desetinné části rychlosti atributu *Rychlost*,
 - d. převedení atributu *TypVozidla10* na nový způsob typologie.

⁸ <https://doprava.plzensky-kraj.cz/opendata/doprava/den/>

4. Ověření správnosti hodnot nových atributů vzhledem ke konkrétnímu účelu.
5. V případě výskytu chyby záznam vyloučit a pokračovat od bodu 1 s novým záznamem, jinak pokračovat dále.
6. Vytvoření a zápis nového záznamu, nebo uchování záznamu v paměti pro další zpracování.

Následuje praktická ukázka popsaného algoritmu na několika záznamech z Obr. 2.

```
10057002|"2016-01-03 00:00:01.099"|1|1.00|100.00|58.00|0|2|0|""|2
10117201|"2016-01-03 00:06:39.025"|1|1.00|100.00|48.00|0|2|0|""|2
10123102|"2016-01-03 07:25:27.243"|1|2.00|100.00|0.00|0|4|9|""|6
```

Obr. 2: Několik ilustračních záznamů z původních dat

Po načtení záznamu je rozdělen dle dělicího znaku svislé čáry. Zároveň dojde k vyřazení nerelevantních atributů, což lze učinit na základě jejich pořadí. Pro ilustrační záznamy zobrazuje výsledek tohoto kroku Tab. 7.

Tab. 7: Rozdělení relevantních dat do skupin podle dělicího znaku

IdDetektor	DatumCas	Rychlost	TypVozidla10
10057002	"2016-01-03 00:00:01.099"	58.00	2
10117201	"2016-01-03 00:06:39.025"	48.00	2
10123102	"2016-01-03 07:25:27.243"	0.00	6

Následujícím krokem je převedení těchto hodnot do nové podoby a zároveň vhodného datového formátu pro daný typ atributu. Podobu ilustračních záznamů demonstruje Tab. 8, kde mohou být jednotlivé atributy zaznamenány v podobě obecného celého čísla, v případě označení směru je také možné zvolit datový typ `boolean`.

Tab. 8: Nové rozdělení dat a úpravy formátu

Id zařízení	Typ záznamu	Směr	Čas	Rychlost	Typ vozidla
057	0	0	00:00:01.099	58	0
117	2	1	00:06:39.025	48	0
123	1	0	07:25:27.243	0	3

Při práci s časovým zařazením je možné využít jak speciální datový formát pro časové hodnoty, tak případný převod na jednotné celé číslo, což může usnadnit a urychlit případné řazení a porovnávání dat.

Dalším krokem je odstranění chybných záznamů s ohledem na to, k jakému účelu mají data dále sloužit. Pokud by se jednalo pouze o počítání průjezdů v obou směrech, pak je možné použít všechny tři záznamy. V případě, že by bylo účelem počítat pouze vozidla ve směru k zařízení, byl by z těchto záznamů použit pouze druhý. Pro určení rychlosti vozidel by byl použit první a druhý záznam. Pro určení typologie vozidel pak pouze třetí.

Posledním možným krokem je pak zapsání výsledných záznamů do souboru. Atribut s časovým zařazením je převeden do vhodnější podoby vzhledem k typu. Obecný formát tohoto záznamu má pak následující podobu:

```
id_zařízení;typ_pořízení;směr;čas;rychlost;typ_záznamu
```

Konkrétní podoba pro ilustrační záznamy je na Obr. 3.

```
57;0;0;00:00:01.099;58;0  
117;2;1;00:06:39.025;48;0  
123;1;0;07:25:27.243;0;4
```

Obr. 3: Ilustrační záznamy v nové podobě

8.2 Využití časové agregace

V této podkapitole je demonstrováno několik způsobů transformace otevřených dat s uplatněním časové agregace, přičemž jako základ pro tyto metody slouží očištění a shromáždění dat z předchozího bodu. Na základě tohoto časového dělení je poté demonstrováno několik dalších způsobů, jak je možné agregaci dat uplatnit.

Pro identifikaci zařízení je v této podkapitole použita hodnota označující dané zařízení, která je dostupná přímo v souboru se záznamy. Není zde proto nutné využívat soubor `Locations`.

8.2.1 Časová agregace s počtem průjezdů

Prvním a zároveň nejvíce obecným způsobem použití agregace je sjednocení záznamů do časových bloků pevné velikosti a poskytnutí výstupní informace v podobě počtu průjezdů vozidel pro každou záznamovou stanicí.

Tento způsob transformace je vhodné použít u typu simulace, která zohledňuje pouze hustotu dopravy, případně v situacích, kdy jsou ostatní relevantní informace známy.

Pro tento způsob transformace je důležitou informací jednoznačná identifikace místa, identifikace směru a časová identifikace záznamu. Případné nepřesnosti v určení rychlosti, nebo typu vozidla není třeba zohledňovat.

Popis algoritmu pro výpočet této agregace:

1. Načtení a očištění jednotlivých záznamů.
2. Časové seřazení záznamů.
3. Vytvoření datových struktur pro jednotlivá zařízení:
 - a. vytvoření a určení hranic časových bloků dle jejich velikosti,
 - b. rozdělení dat podle směrů,
 - c. vytvoření hodnoty pro počítání průjezdů.
4. Nastavení prvního časového bloku a jeho horní hranice.
5. Postupný průchod načtenými záznamy:
 - a. porovnání časové hodnoty záznamu s horní mezí současného časového bloku a případné posunutí na blok následující,
 - b. započtení záznamu k příslušnému zařízení a směru.
6. Vytvoření výstupního souboru v patřičném formátu.

Následuje demonstrace tohoto algoritmu na ilustrační množině dat, kterou jsou v tomto případě průjezdy vozidel pro stanici s označením 51. Velikost časového bloku jsou v tomto ilustračním příkladu tři hodiny.

Příklad několika záznamů z dané stanice demonstruje Obr. 4.

```
10051001|"2016-01-03 00:00:38.079"|1|1.00|100.00|49.00|0|2|0|""|2  
(...)  
10051001|"2016-01-03 02:59:44.897"|1|1.00|100.00|48.00|0|2|0|""|2  
10051002|"2016-01-03 03:06:18.545"|1|1.00|100.00|86.00|0|2|0|""|2  
(...)
```

Obr. 4: Několik původních záznamů pro ilustraci transformace

Po načtení těchto záznamů dojde k jejich očištění o nerelevantní data. V tomto případě jsou dále použita pouze data o místě pořízení záznamu, směru a časovém zařazení.

Dále jsou vytvořeny datové struktury pro načtení záznamů, což může být dvourozměrné pole, jehož první rozměr je počet časových bloků, zde v příkladu osm, a jehož druhý rozměr je počet směrů, zde dva.

Následuje postupný průchod záznamy a jejich započtení k patřičnému zařízení ve správném časovém bloku. Vzhledem k tomu, že data byla nejdříve seřazena podle času, je možné individuální záznamy zařazovat pouze na základě porovnání s maximální časovou hodnotou současného časového bloku.

Na začátku je časový blok nastaven na hodnotu 0 a jeho maximální hodnota je 2:59:59.999. Jak algoritmus postupuje záznamy, dochází k jejich porovnání s touto hodnotou. V případě prvních dvou ilustračních záznamů tak dojde k jejich zařazení do prvního časového bloku. Ve chvíli, kdy přijde na řadu třetí záznam, jeho časová hodnota překročí horní mez prvního bloku a dojde tedy k navýšení bloku na hodnotu 1 a posunutí maximální hodnoty na 5:59:59.999. Dojde opět k porovnání s maximální hodnotou současného bloku. V tuto chvíli již záznam vyhovuje a je započten do druhého bloku. Vzhledem k časovému seřazení dat se jedná o bezpečný postup.

Nakonec je vytvořen výstupní formát pro tato vypočtená data. Tato data jsou řazena nejprve podle identifikace zařízení, poté podle označení časového bloku a nakonec podle směru.

Obecný formát výstupu této transformace lze zapsat takto:

```
id_zařízení;směr;časový_blok;počet_průjezdů
```

Pouze pro ilustrační záznamy, při vynechání přeskočených záznamů, výsledek demonstruje Obr. 5.

```
51;1;0;2  
51;0;0;0  
51;1;1;0  
51;0;1;1  
(...)  
51;1;7;0  
51;0;7;0
```

Obr. 5: Demonstrace výsledného formátu časové agregace pro ilustrační záznamy

V případě agregace všech záznamů pro stanici 51 výsledek demonstruje Obr. 6.

```
51;1;0;220  
51;0;0;128  
51;1;1;883  
51;0;1;717  
(...)  
51;1;7;887  
51;0;7;458
```

Obr. 6: Demonstrace výsledného formátu časové agregace pro všechny záznamy

Další demonstrace tohoto způsobu transformace a toho, jak výsledky vypadají při aplikaci na data z jednoho dne ze všech stanic, jsou dostupné na příloženém CD.

8.2.2 Časová agregace s průměrnou rychlostí

Dalším způsobem časové agregace dat je taková, která kromě počtu průjezdů zohledňuje také průměrnou rychlost vozidel, která měřící stanicí během časového bloku projela.

Tento způsob agregace je možné uplatit u těch typů simulace dopravy, které kromě hustoty dopravy v daném místě blíže zohledňují také rychlosti vozidel a zároveň tato data neodvozují pouze podle rychlostního omezení silnice, neboť by takový přístup nemusel odpovídat realitě. Příkladem může být komunikace, na které je dle zákona možno jet až 90 km/h, ale vzhledem k technickému stavu vozovky se vozidla běžně pohybují rychlostí 50 km/h.

V tomto případě je kromě velikosti časových bloků také třeba zohlednit, o jaký typ záznamu se jedná, neboť záznamy pořízené pomocí kamery neposkytují informace o průjezdech v protisměru, tudíž ani informace o rychlostech vozidel. Tento problém se navíc dále přenáší k záznamům pořízených video detekcí, která není schopná sama o sobě rychlost vozidel určit a používá tak některý ze zbylých dvou způsobů detekce pro její určení, což znamená v případě snahy určit rychlost vozidel v protisměru kamerou nulovou hodnotu. Informace o tom, který typ byl použit, se v datech nenachází.

Jak výsledné hodnoty pro různé typy záznamů vypadají bez použití patričného filtrování, demonstruje Tab. 9. Hodnoty jsou v tomto příkladu agregovány za celý den.

Tab. 9: Průměrné rychlosti podle typu zařízení bez filtrování; st. č. 51

	ve směru	v protisměru
radarový záznam	46	55
video záznam	43	29
kamerový záznam	47	-

Pro tento způsob transformace je možné využít algoritmus z předchozího bodu s jistými úpravami:

1. Načtení a očištění záznamů.
2. Časové seřazení záznamů.

3. Vytvoření datových struktur pro jednotlivá zařízení:
 - a. vytvoření a určení hranic časových bloků dle jejich velikosti,
 - b. rozdělení dat podle směrů,
 - c. vytvoření hodnoty pro počítání průjezdů,
 - d. vytvoření hodnoty pro čítání rychlosti a čítání relevantních průjezdů.
4. Nastavení prvního časového bloku a jeho horní hranice.
5. Postupný průchod načtenými záznamy:
 - a. porovnání časové hodnoty záznamu s horní mezí současného časového bloku a případné posunutí na blok následující,
 - b. započtení záznamu k příslušnému zařízení a směru,
 - c. kontrola typu pořízení a případné započtení informace o rychlosti.
6. Vypočtení průměrné rychlosti za časový blok.
7. Vytvoření výstupního souboru v patřičném formátu.

Při demonstraci této transformace je možné využít stejný vzorek dat, jako v předchozím bodě.

V případě této transformace je však množina relevantních dat větší, než v předchozím. Kromě dat o času, místě a směru je také nutné pracovat s informací o typu pořízení záznamu a hodnotou rychlosti průjezdu.

Datová struktura pro uchování dat o průjezdech daným zařízením může mít totožný formát, jako v předchozím případě. Uchování údajů o rychlosti pak může mít obdobnou strukturu formátu, přičemž uchovávanou hodnotou může být součet rychlostí, nebo postupně vypočítávaná průměrná rychlost. Z tohoto důvodu je také nutné uchovávat informaci o tom, kolik průjezdů je pro určení průměrné rychlosti relevantních.

Následně dojde postupnému průchodu všech záznamů, jejich časovému zařazení - stejně, jako v předchozím případě -, ale také započtení údajů o rychlosti.

Nakonec je třeba informace o rychlostech průjezdů převést do podoby průměrné rychlosti, získaná data převést do podoby výsledného formátu a zapsat do souboru.

Nové záznamy mají následující obecný formát:

```
id_zařízení;směr;časový_blok;průměrná_rychlost;počet_průjezdů
```

Při časové agregaci za tři hodiny pak záznamy vydají pro stanici č. 51 jako na Obr. 7.

```
51;1;0;51.03;220
51;0;0;62.08;128
51;1;1;47.72;883
51;0;1;52.9;717
```

(...)
51; 1; 7; 46.7; 887
51; 0; 7; 56.14; 458

Obr. 7: Demonstrace výsledného formátu časové agregace s průměrnou rychlostí

Podrobnější příklady výsledku této transformace pro data ze všech stanic za jeden den je možné nalézt na příloženém CD.

8.2.3 Časová agregace s typologií vozidel

Dalším způsobem transformace, který poskytuje množinu dat navíc, je využití členění vozidel dle typů.

Tento způsob transformace je možné využít u těch typů simulace silniční dopravy, kde informace o tom, o jaký typ vozidla se jedná, významně rozhoduje o jeho dalším chování v systému a o jeho případných omezeních. Konkrétněji se může jednat o rozdílné vlastnosti zrychlení, schopnosti zatáčet, nebo předjíždět, ale také o možné omezení maximální rychlosti, nebo omezení míst, kam se může dané vozidlo v rámci simulace vydat.

V případě této simulace je rovněž třeba zohlednit typ pořízení záznamu, neboť pouze video záznam spolehlivě určuje, o jaký typ vozidla se jedná. V případě zbylých typů pořízení záznamu se jedná o hodnoty osobní automobil, ale ty je možné považovat za chybné, respektive nepřesné, a je proto vhodné zařadit je do této kategorie ve výsledných datech.

Dříve popsaný algoritmus má tak v tomto případě následující podobu:

1. Načtení a očištění záznamů.
2. Časové seřazení záznamů.
3. Vytvoření datových struktur pro jednotlivá zařízení:
 - a. vytvoření a určení hranic časových bloků dle jejich velikosti,
 - b. rozdělení dat podle směrů,
 - c. vytvoření hodnot pro počítání průjezdů pro všechny typy vozidel,
 - d. vytvoření hodnoty pro čítání rychlosti a čítání relevantních průjezdů.
4. Nastavení prvního časového bloku a jeho horní hranice.
5. Postupný průchod načtenými záznamy:
 - a. porovnání časové hodnoty záznamu s horní mezí současného časového bloku a případné posunutí na blok následující,

- b. kontrola typu pořízení a započtení záznamu ke správnému typu vozidla u příslušného zařízení a směru,
 - c. kontrola typu pořízení a případné započtení informace o rychlosti.
6. Vypočtení průměrné rychlosti za časový blok.
 7. Vytvoření výstupního souboru v patričném formátu.

Mezi relevantní data v tomto případě patří označení zařízení a směru, typ pořízení záznamu, čas průjezdu, rychlost a nově také typ vozidla.

Vlastní postup je pak téměř totožný jako v předchozím případě s tím rozdílem, že pokud se nejedná o záznam pořízení videem, je vyhodnocen jako nepřesný pro určení typu, v opačném případě je převedena původní hodnota na novou a započítána.

Výsledný formát je rovněž téměř shodný, avšak hodnota pro počty průjezdů za daný časový úsek není celé číslo, ale je to pole celých čísel, kde jeho pořadí určuje typ vozidla.

Obecný formát pro tuto transformaci má následující podobu:

```
id_zařízení;směr;čas;prům_rychlost;[typ0;typ1;typ2;typ3;typ4]
```

Jako ilustrační příklad výsledků je opět použita stanice č. 51 s agregací hodnot po třech hodinách, na Obr. 8.

```
51;1;0;51.03;[142;78;0;0;0]
51;0;0;62.08;[112;14;0;0;0]
(...)
51;1;3;46.28;[2124;365;22;71;55]
51;0;3;51.8;[990;338;16;20;24]
(...)
51;1;7;46.7;[577;310;0;0;0]
51;0;7;56.14;[373;55;0;0;0]
```

Obr. 8: Demonstrace výsledného formátu časové agregace s typologií vozidel

Podrobněji je tato transformace demonstrována na příloženém CD.

8.2.4 Časová agregace dělená podle typů záznamu

Jak bylo popsáno dříve, jednotlivé typy pořízení záznamů poskytují rozdílné informace s různou přesností. Řešením uplatněným v předchozích transformacích bylo toto omezení zohlednit a data patričně filtrovat. Dalším možným přístupem, popsaným v tomto bodě, je tato typová omezení neřešit během transformace přímo, ale místo toho výsledná data rozdělit na základě typů pořízení.

Tento typ je, obdobně jako pouhé očištění záznamů, vhodný v případech, kdy je dále kombinován s dalšími transformacemi, případně je možné jej využít při analýze jednotlivých typů pořízení, spíše než jako praktická vstupní data pro simulaci.

Algoritmus pro tuto transformaci vypadá následovně:

1. Načtení a očištění záznamů.
2. Časové seřazení záznamů.
3. Vytvoření datových struktur pro jednotlivá zařízení:
 - a. vytvoření a určení hranic časových bloků dle jejich velikosti,
 - b. rozdělení dat podle směrů,
 - c. rozdělení dat podle typu pořízení,
 - d. vytvoření hodnoty pro počítání průjezdů.
4. Nastavení prvního časového bloku a jeho horní hranice.
5. Postupný průchod načtenými záznamy:
 - a. porovnání časové hodnoty záznamu s horní mezí současného časového bloku a případné posunutí na blok následující,
 - b. započtení záznamu k příslušnému zařízení, směru a typu záznamu.
6. Vytvoření výstupního souboru v patřičném formátu.

Vlastní algoritmus je v tomto případě uplatněn obdobně, jako v případě, kdy je účelem zjištění počtu průjezdů, avšak tato data jsou u daného zařízení dělena nejen podle směru průjezdu, ale také podle typu pořizovacího zařízení.

Obecný formát vypadá následovně:

```
id_zařízení;typ_záznamu;směr;čas_blok;počet_průjezdů
```

Nové záznamy pro stanici č. 51, pro demonstrační účely použit časový blok 24 hodin, zobrazuje Obr. 9.

```
51;0;1;0;6267  
51;0;0;0;6162  
51;1;1;0;3727  
51;1;0;0;3157  
51;2;1;0;4819  
51;2;0;0;0
```

Obr. 9: Demonstrace výsledného formátu časové agregace s typy záznamů

8.3 Využití poměrových ukazatelů

Vzhledem ke způsobu, jakým je v otevřených datech zaznamenána typologie vozidel, je pro získání realistického pohledu na dopravu nezbytné tyto hodnoty transformovat složitějším způsobem, než je pouhá agregace záznamů.

Jedno z možných řešení, členění záznamů podle typu jejich pořízení, demonstroval předchozí bod.

Dalším řešením, které již toto členění nevyužívá přímo, je použít pro typologii vozidel pouze relevantní typ záznamů a výsledné hodnoty poté vyjádřit jako poměry jednotlivých typů vozidel vůči relevantním záznamům. Tyto hodnoty jsou dále doplněny o celkový počet průjezdů, z čehož je možné roznásobením získat předpokládané absolutní hodnoty průjezdů, nebo jsou tyto hodnoty rovnou přepočítány.

8.3.1 Časová agregace s poměry typů vozidel

Konkrétním případem tohoto způsobu transformace je použití časových bloků v kombinaci s počítáním celkového počtu průjezdů a poměrového vyjádření typů vozidel z dat získaných video detekcí.

Hodnoty průjezdů jsou počítány ze všech typů průjezdů a jsou následně doplněny polem hodnot, které jsou ze záznamů pořízených pomocí video detekce vypočítány jako poměr daného typu vozidel vůči všem relevantním hodnotám.

Popis algoritmu:

1. Načtení a očištění záznamů.
2. Časové seřazení záznamů.
3. Vytvoření datových struktur pro jednotlivá zařízení:
 - a. vytvoření a určení hranic časových bloků dle jejich velikosti,
 - b. rozdělení dat podle směrů,
 - c. vytvoření hodnoty pro počítání průjezdů,
 - d. vytvoření hodnot pro čítání typů a všech relevantních průjezdů.
4. Nastavení prvního časového bloku a jeho horní hranice.
5. Postupný průchod načtenými záznamy:
 - a. porovnání časové hodnoty záznamu s horní mezí současného časového bloku a případné posunutí na blok následující,
 - b. započtení záznamu k příslušnému zařízení a směru,

- c. kontrola typu pořízení a případné započtení typu i záznamu.
- 6. Vypočtení poměrů jednotlivých typů vozidel.
- 7. Vytvoření výstupního souboru v patřičném formátu.

Postup vlastní transformace je obdobný jako v předchozích případech, proto není třeba jej blíže, krok po kroku, rozepisovat.

Důležitým rozdílem je počet záznamů v poli s typy vozidel. Vzhledem ke způsobu, jak jsou v této transformaci tyto záznamy hodnoceny, není nutné používat speciální záznam pro chybnou a nejasnou hodnotu. Pole s typy vozidel má v tomto případě čtyři hodnoty, přičemž chybový typ, první v pořadí, je vynechán.

Výsledný formát dat má následující obecnou podobu:

```
id_zařízení;směr;časový_blok;počet_průjezdů;[typ1;typ2;typ3;typ4]
```

Při agregaci za celý den pak tyto hodnoty pro stanici 51 demonstruje Obr. 10.

```
51;1;0;14813;[0.88087;0.01690;0.06010;0.04213]
51;0;0;9319;[0.90117;0.02376;0.04435;0.03073]
```

Obr. 10: Demonstrace výsledného formátu při využití poměrových ukazatelů typů

8.3.2 Časová agregace s přepočtenými poměry vozidel

Další aplikací předchozího přístupu je vypuštění vypočtených poměrů z dat a zahrnutí přepočtených čísel v nových záznamech rovnou v podobě průjezdů vozidel.

Tento přístup lze zároveň považovat za eliminaci chyb měření a je tedy možné jej použít u těch typů simulace, kde je typ vozidla relevantní informací.

V tomto případě je použit stejný algoritmus, jako v předchozím případě, ale před vypsáním dat dojde k patřičnému roznásobení poměrových hodnot celkovým počtem průjezdů vozidel.

Formát výstupu má pak následující podobu, kdy jednotlivé typy reprezentují typologii vozidel:

```
id_zařízení;směr;časový_blok;[typ1;typ2;typ3;typ4]
```

Provedení této transformace v praxi demonstrují následující záznamy:

```
51;1;0;[13048;250;890;624]
51;0;0;[8398;221;413;286]
```

Obr. 11: Demonstrace výsledného formátu při přepočtení poměrových ukazatelů typů

9 Zhodnocení výsledků

Tato kapitola se zabývá vyhodnocením provedené analýzy a transformací se zaměřením na vyhodnocení rizik, se kterými bylo třeba se během analýzy a transformací potýkat.

9.1 Zhodnocení analýzy dat

Následující body popisují faktory, které mohou ovlivnit podobu vstupních dat, a navrhuje, nebo nastiňuje možná řešení.

9.1.1 Nefunkční stanice

Ačkoliv se na webových stránkách dopravy plzeňského kraje, případně stránkách výrobce zařízení, nachází také přehled informující o tom, která zařízení jsou momentálně mimo provoz, jedná se zejména o přehled těch zařízení, u kterých došlo k dlouhodobé a plánované odstávce. V případě krátkodobých výpadků, nebo prosté nefunkčnosti těchto zařízení nejsou tyto informace dostupné online ani jako součást metadat.

Případná nefunkční zařízení je tak třeba identifikovat na základě chybějících záznamů za dané časové období, ačkoliv je zařízení zmíněno v souboru `Locations`.

Dalším krokem je pak případná identifikace zařízení s výpadky, nebo chybami měření (např. z důvodu rozbitého čidla). Jedná se o zařízení, která za měřený úsek zaznamenala neúměrné množství záznamů oproti běžným situacím.

V obou zde zmíněných případech je vhodná další analýza s použitím dalších zdrojů informací, neboť důvodem chybějících, nebo zkreslených záznamů nemusí být nutně vada na zařízení, ale kupříkladu dočasná uzavírka, nebo omezení provozu po dané komunikaci.

9.1.2 Nefunkční měřicí zařízení

Další problém, který se může vyskytnout, je případ, kdy u měřicího zařízení nefunguje jeden, nebo dva způsoby sběru dat.

Tento problém je ve většině případů snadné identifikovat, neboť pro kontrolu stačí u jednotlivých zařízení průběžně počítat, z jakého typu měřicího zařízení byl záznam pořízen. Za předpokladu, že se tyto hodnoty nevyskytují, lze předpokládat, že daný typ měřicího zařízení u stanice nefunguje.

Na základě této informace je možné vyvodit důsledky pro všechna data z takové stanice. Například v případě chybějícího radaru nedojde k realistickému zaznamenávání rychlostí průjezdu, v případě chybějící video detekce pak není možné pro danou stanici určit typologii vozidel. Jakékoliv chybějící měřící zařízení navíc znamená zkreslení celkové podoby dopravy, neboť některé průjezdy nemusí být zaznamenány.

Tab. 10 popisuje procentuální rozložení dopravy v rámci agregace za celý den s rozdělením podle typu měřících zařízení v rámci několika dnů z různých časových období. Rozsáhlejší data jsou dostupná v příloze *D – Ostatní* na přiloženém CD v podobě poměrů i pouhých součtů. Na základě obdobných informací je možné vytvořit intervaly běžných hodnot a hodnot, které je možné vyhodnotit jako chybné, nebo podezřelé a to například vzhledem ke všem datům, nebo datům za několik předcházejících a následujících dnů. Podezřelé dny mohou být následně reportovány pro ruční revizi, nebo vyřazovány.

Tab. 10: Ilustrační příklad procentuálního rozložení záznamů podle typu pořízení

	Radarový záznam	Video záznam	Kamerový záznam
<i>1. 1. 2016</i>	55,41 %	29,81 %	14,78 %
<i>5. 6. 2017</i>	46,79 %	36,04 %	17,17 %
<i>29. 10. 2018</i>	49,42 %	37,58 %	13 %

9.1.3 Změna podoby dat

V rámci této práce byla prozkoumávána a analyzována data od počátku roku 2016, do prvního čtvrtletí roku 2019. Během této doby došlo v datech k několika drobným změnám, které bylo třeba při tvorbě transformačních algoritmů zohlednit a lze předpokládat, že takovéto změny se mohou objevit i v budoucnu.

Kupříkladu v případě dřívějších souborů byly záznamy průjezdů řazeny přirozeně od prvního průjezdu dne, do posledního a nebylo tak nutné je řadit dále. V pozdějších archivech pak byla data nejdříve řazena podle identifikačního čísla dané stanice a až poté podle času záznamu. Tyto způsoby se později ještě několikrát mění a dochází i k případům, kdy se mezi záznamy objevují některé špatně seřazené.

Na základě vyhodnocení prozatím dostupných dat také bylo možné zredukovat typy vozidel na kategorie, které jsou skutečně využity, což se může v budoucnu rovněž měnit.

Dalším případem, který se však netýká přímo dat z portálu plzeňského kraje, ale týká se rovněž projektu podobného portálu od firmy GEMOS, konkrétně toho ve Středočeském kraji, je výskyt nových typů zařízení a nových souborů. Středočeská data mají v novějších souborech dodatečné informace z vah, které jsou nainstalovány u některých stanic. Tyto stanice poskytují jinou množinu dat v separátním souboru a je využita jiná konstrukce cizích klíčů. Ačkoliv se nejedná o data z plzeňského kraje, je zde možnost, že při instalaci případných nových zařízení s touto funkcionalitou budou tato data dostupná i zde.

9.1.4 Chybějící informace

Ačkoliv by záznamové stanice měly být, dle deklaráce výrobce, schopné pořizovat také informace o tom, v jakém jízdním pruhu se vozidlo pohybovalo, nejsou tato data v otevřených datech k dispozici.

Tato informace by mohla být hodnotná zejména v případech, kdy se měřící zařízení nachází v bezprostřední blízkosti křižovatky. Na základě informací o pruhu, ve kterém se vozidlo pohybovalo, by pak bylo možné kupříkladu vytvořit nové generační místo, které by stanici dále dělilo na průjezdy podle případného odbočení.

Další potenciálně důležitou informací, kterou ze současných dat není možné získat, je poměr vozidel, které projely více zařízeními vzhledem k jejich relativní blízkosti a výskytu na významných komunikacích. Tento jev tak vede k jistému nabobtnání dopravy v dané oblasti.

9.1.5 Možný vícenásobný záznam

V některých případech je možné si všimnout, že bylo pořízeno několik záznamů různých typů v krátkém časovém sledu (0,5-1 vteřina). Jedním z možných vysvětlení je, že někdy dochází k pořízení více záznamů jednoho vozidla, druhým vysvětlením je, že se v případě hustých průjezdů zařízení střídají.

Pro vyhodnocení tohoto jevu v datech je potřeba další, hlubší analýzy, kde je nutné využít alternativní zdroje (např. ruční sběr dat, nebo pořízení vlastního video záznamu v souladu s legislativou).

V prvním případě by tak docházelo ke zkreslení počtu průjezdů, případně také zkreslení dat o typech vozidel. Ve druhém případě dochází ke zkreslení dat jen omezením typu záznamu. Možným řešením je data ze stejné stanice, ve stejném směru za určitý krátký časový úsek sjednocovat. Druhým je použití pouze jednoho typu záznamů. To však není vhodné, neboť ve většině případů je pořízen záznam pouze jednoho z typů.

9.2 Zhodnocení užitých transformací

Tato podkapitola se zaměřuje na vyhodnocení informační hodnoty nové podoby záznamů oproti původním záznamům. Dále na to, jaké nepřesnosti a případné ztráty informací se mohou v těchto nových záznamech vyskytnout v důsledku úpravy jejich podoby a to jak v případě nových jednotlivých záznamů, tak v případě datové agregace. Nakonec je vyhodnocena výpočetní složitost použitého algoritmu.

9.2.1 Informační hodnota nových záznamů

Původní data obsahují několik atributů, které neposkytují pro simulaci silniční dopravy žádné relevantní informace. Jedná se o data, která jsou sbírána jen u některých typů záznamů a pravděpodobně slouží pro pomocné výpočty v rámci zařízení.

Původní data také obsahují atributy typologie, jejichž informace je v datech redundantní, neboť se v datech vyskytuje celkem třikrát. Ve zjednodušené podobě, v rozsáhlejší dělení a poté v atributu `IntenzitaN`.

Z tohoto důvodu bylo možné nepotřebné kategorie ze záznamů vyřadit bez ztráty informací a z atributů pro typologii vozidel vybrat ten, který má nejvhodnější hodnoty.

Na straně druhé, v případě identifikace měřicí stanice obsahuje tento atribut hned několik různých informací. V tomto případě tedy bylo vhodným krokem tyto hodnoty rozdělit do separátních atributů, na jejichž základě mohou být záznamy dále posuzovány, nebo filtrovány.

9.2.2 Ztrátovost při agregaci

V rámci popsaných transformací dochází k agregacím a přepočtům několika hodnot, což může vést ke ztrátě přesnosti dat.

V první řadě se jedná o agregaci jednotlivých průjezdů do kumulativních hodnot za časová období. Tato agregace je bezztrátová za předpokladu, že je pro uchování těchto

hodnot použít datový typ vhodné velikosti, neboť celkový počet průjezdů zůstává stále zachován.

Další agregací je výpočet průměrné rychlosti z rychlostí uvedených v jednotlivých průjezdech. V tomto případě vycházíme z dat, která jsou již do jisté míry zkrácena, neboť jsou v rámci otevřených dat poskytována zaokrouhlena na celé kilometry za hodinu. Při výpočtu průměrné rychlosti pak při využití hodnot s přesností na několik desetinných míst dochází k zanedbatelné ztrátě. Důvodem pro případné nevyužití informací je tak spíše odchylka, která vznikla při vlastním měření rychlosti, případně následném zaokrouhlení na celé číslo.

Poslední agregovanou hodnotou je typologie vozidel, kde vzhledem k různým způsobům záznamů dochází k nepřesnostem již ve vstupních datech. Pokud jsou data z nevhodných záznamů vyhodnocena jako nepřesná a podle toho náležitě zařazena do kategorie, pak dochází ke ztrátě pouze v případě, kdy by u součtů došlo k přetečení z důvodu nevhodně zvolené datové struktury. V případě využití poměrových ukazatelů k přepočtení typologie vozidel může docházet ke zkrácení z důvodu, že video záznam tvoří pouze zhruba 31 % všech záznamů (jak demonstruje Tab. 10).

9.2.3 Vyhodnocení složitosti algoritmu

Každý záznam je nejprve nutné načíst a vhodným způsobem parsovat, což je usnadněno formátem CSV a konzistentní podobou atributů. Nerelevantní atributy mohou být přeskočeny, zbylé atributy jsou následně převedeny na vhodný datový typ.

U identifikace zařízení je atribut dále rozdělen na tři hodnoty. Toto parsování je usnadněno jednotným formátem tohoto záznamu, není proto třeba dalších kontrol a pomocných operací.

Výše zmíněné kroky jsou provedeny stejným způsobem pro všechny záznamy.

Vzhledem k malému množství záznamových stanic a faktu, že identifikací stanice je celé, maximálně trojčíferné číslo, je možné využít jako datovou strukturu pro uložení informací o dané stanici tabulku s přímým adresováním. Výhodou této datové struktury je možnost přidávat a vyhledávat hodnoty se složitostí $O(1)$. Jednotlivé záznamy mohou být následně vloženy na konec spojového seznamu se složitostí rovněž $O(1)$. Načtení záznamů a jejich uložení je tedy možné se složitostí $O(n)$, kde n je počet záznamů.

Následně je třeba zvolit vhodný algoritmus pro seřazení dat. Zde je třeba zohlednit fakt, že v některých případech jsou tato data již seřazena. Například při použití algoritmu *heapsort* dojde k seřazení se složitostí $O(n \cdot \log(n))$.

Seřazená data je již možné využít pro vypočítání agregací. Je dopředu známo, na kolik částí se budou výsledná data dělit a je proto možné datové struktury navrhnout opět tak, aby k nim bylo možné přistupovat se složitostí $O(1)$.

Při agregaci dochází k postupnému procházení záznamů jednotlivých stanic. Dojde k porovnání s maximální hodnotou současného časového intervalu a případnému navýšení této hodnoty. Následně je záznam přičten do čítače průjezdů. Poté dojde na základě typu záznamu k přičtení hodnoty průměrné rychlosti, nebo typu vozidla, přičemž při typovém rozdělení je vztah následující $n=r+v+k$, kde n je počet všech průjezdů, r jsou radarové záznamy, v jsou video záznamy a k jsou kamerové záznamy. Operace porovnání času a započtení průjezdu proběhne vždy n -krát, porovnání typu záznamu až $2n$ -krát. Operace započtení rychlosti proběhne r -krát a operace započtení typu pak v -krát. V nejlepším případě pak proběhne $4n$ operací, v nejhorším případě pak $5n$ operací, pokud se r , nebo v rovná n a zároveň je to až druhá porovnávaná hodnota.

Operace pro výpočet finálních hodnot pro jednotlivé datové bloky pak proběhne v závislosti na počtu těchto časových bloků, stejně jako případné navýšení časového bloku při procházení záznamů. Výsledné hodnoty je poté třeba zapsat do výstupního souboru.

10 Závěr

Tato práce poskytuje dostatečný teoretický základ pro porozumění otevřeným datům poskytovaným v rámci projektu Klidné příhraničí na stránkách dopravy Plzeňského kraje v kontextu simulace silniční dopravy. Dále detailně popisuje význam jednotlivých atributů těchto dat a formát jejich uložení.

Následně navrhuje několik způsobů, jak je možné tato data převést z původní podoby do podoby, kdy mohou být využita jako vstupní data v simulaci silniční dopravy s ohledem na různou úroveň detailu, případně na lišící se množinu vstupních dat.

Zmíněné transformace jsou dále demonstrovány na testovací množině dat s patřičným popisem algoritmů. Na tyto transformace je možno dále navázat případnými úpravami popsaných algoritmů, nebo volbou jiné množiny, nebo podoby výstupních dat v závislosti na konkrétním typu simulace a užitém modelu. Také je možné blíže analyzovat problémy, které tato práce identifikuje, případně popisuje, ale neřeší je.

Další využití by tato práce mohla nalézt také u obdobného projektu, který rovněž poskytuje otevřená data v takovéto podobě a v současné době probíhá ve Středočeském kraji. Je proto možnost aplikovat zde popsané algoritmy i na data z takového projektu, nebo jiných projektů se zařízeními firmy GEMOS. V případě jiných projektů, jejichž součástí jsou rovněž otevřená data, je možné použít podobnou analýzu, upravenou na míru, a na jejím základě vytvořit vlastní algoritmy a transformace.

Seznam literatury

- [1] J.S. Kaizer, A.K. Heller a W.L. Oberkampff: „Scientific computer simulation review“, *Reliability Engineering and System Safety*, svazek 138, pp. 210–218, 2015
- [2] R.M. Fujimoto: *Parallel and Distributed Simulation Systems*, John Wiley & Sons, 2000
- [3] T.B. Brito, R.C. Botter a E.F.C. Trevisan: „A conceptual comparison between discrete and continuous simulation to motivate the hybrid simulation methodology“, *Proceedings of the 2011 Winter Simulation Conference*, Piscataway, New Jersey: IEEE, 2011, pp. 3910-3922
- [4] R.G. Ingalls: „Introduction to simulation“, *Proceedings of the 2008 Winter Simulation Conference*, San Diego, California, 2002, pp. 7-16
- [5] A.I. Delis, I.K. Nikolos, M. Papageorgiou: “Macroscopic Modelling and Simulation of Multi-Lane Traffic”, *IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, Spain, 2015, pp. 2213-2218
- [6] D. Ni: „2DSIM: A Prototype of Nanoscopic Traffic Simulation“, *Intelligent Vehicles Symposium. Proceedings*, Columbus, Ohio, 2003, pp. 47-52
- [7] S.O. Simonsson: „Car-Following as a Tool in Road Traffic Simulation“, *Proceedings of the IEEE-IEE Vehicle Navigation and Information Systems Conference*, Ottawa, Ontario, Canada, 1993, pp. 150-153
- [8] Y. Tian a Y.-C. Chiu: „Anisotropic Mesoscopic traffic Simulation approach to support large-scale traffic and logistic modeling and analysis“, *Proceedings of the 2011 Winter Simulation Conference*, Phoenix, Arizona, 2012, pp. 1495-1507
- [9] P.B. Kotagi a G. Asaithambi: “Simulation Framework for Modeling Bidirectional Mixed Traffic”, *Communication Systems and Networks (COMSNETS)*, Bangalore, India, 2017, pp. 443-447
- [10] O. Özgün a Y. Barlas: „Discrete vs. Continuous Simulation: When Does It Matter?“, *Proceedings of the 27th International Conference of The System Dynamics Society*, Albuquerque, New Mexico, 2009
- [11] J. Harri, F. Filali a C. Bonnet: „Mobility Models for Vehicular Ad Hoc Networks: A Survey and Taxonomy“, *IEEE Communications Surveys & Tutorials*, svazek 11, vydání 4, pp. 19-41, 2009
- [12] D.Wilkie, J. Sewall a M.C. Lin: „Transforming GIS Data into Functional Road Models for Large-Scale Traffic Simulation“, *IEEE Transactions on Visualization and Computer Graphics*, svazek 18, vydání 6, pp. 890-901, 2011

- [13] T.Potužák: *Methods for Reduction of Interprocess Communication in Distributed Simulation of Road Traffic*, Plzeň, 2009, Dizertační práce, Západočeská univerzita v Plzni
- [14] H. Guzmán, M. Lárraga, L. Alvarez-Icaza, F. Huerta: „A Realistic Two-lanes Traffic Simulation Model Based on Cellular Automata“, *Modelling Symposium (EMS)*, Pisa, Italy, 2014, pp. 126-131
- [16] C. Tao, J. Zhang: „A Cellular Automata Simulation on Multi-lane Traffic Flow for Designing Effective Rules“, *2015 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration*, Wuhan, China, 2015, pp. 209-212
- [17] Ředitelství silnic a dálnic ČR, ŘSD ČR – *Mapy* [online], aktualizováno k 1.7.2017 [citováno 18.12.2017]. Dostupné z: <https://www.rsd.cz/wps/portal/web/Silnice-a-dalnice/mapy>
- [18] G. Palubinskas, F. Kurz a P. Reinartz: „Detection of Traffic Congestion in Optical Remote Sensing Imagery“, *Geoscience and Remote Sensing Symposium*, IEEE International, Boston, Massachusetts, 2008, pp. 426-429
- [19] M. Yin et al.: “A Generative Model of Urban Activities from Cellular Data”, *IEEE Transactions on Intelligent Transportation Systems*, vydání 99, pp. 1-15, 2017
- [20] M. Alam et al.: „Smart Cameras Are Making Our Beaches Safer: A 5G-Envisioned Distributed Architecture for Safe, Connected Coastal Areas“, *IEEE Vehicular Technology Magazine*, svazek 12, vydání 4, prosinec 2017, pp. 50-59, 2017
- [21] M. Behrisch, L. Bieker, J. Erdmann a D. Krajzewicz: „SUMO – Simulation of Urban Mobility: An Overview“, *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*, Barcelona, 2011, pp. 55-60
- [22] OpenStreetMap. *OpenStreetMap* [online]. [cit. 1. 3. 2019]. Dostupné z: <https://www.openstreetmap.org>
- [23] Česká komunita OpenStreetMap. *OpenStreetMap.cz* [online]. [cit. 3. 3. 2019]. Dostupné z: <https://openstreetmap.cz>
- [24] MAFRA, a.s. *Mapy.iDNES.cz* [online]. [cit. 1. 3. 2019]. Dostupné z: <http://mapy.idnes.cz/>
- [25] Seznam.cz, a.s. *Mapy.cz* [online]. [cit. 1. 3. 2019]. Dostupné z: <https://mapy.cz>
- [26] Mapy Google. *Google* [online]. [cit. 1. 3. 2019]. Dostupné z: <https://www.google.cz/maps>
- [27] Microsoft. *Bing Mapy* [online]. [cit. 1. 3. 2019]. Dostupné z: <https://www.bing.com/maps>

- [28] Ředitelství silnic a dálnic ČR. *Dopravní informace* [online]. [cit. 3. 3. 2019]. Dostupné z: <http://www.dopravniinfo.cz>
- [29] Správa inf. technologií města Plzně. *Doprava* [online]. [cit. 3. 3. 2019]. Dostupné z: <https://gis.plzen.eu/doprava/>
- [30] Dopravní portál. *GEMOS CZ spol. s r.o.* [online]. [cit. 8. 3. 2019]. Dostupné z: <http://www.gemos.cz/doprava/>
- [31] Středočeský kraj. *Dopravní portál* [online]. [cit. 9. 3. 2019]. Dostupné z: <https://doprava.kr-stredocesky.cz>
- [32] Plzeňský kraj. *Informační systém Plzeňského kraje* [online]. [cit. 9. 3. 2019]. Dostupné z: <https://doprava.plzensky-kraj.cz>
- [33] Shafranovich, Y. *Common Format and MIME Type for CSV Files* [online]. IETF, říjen 2005 [cit. 10. 3. 2019]. Dostupné z: <https://www.ietf.org/rfc/rfc4180.txt>
- [34] P.Bureš, O. Příbyl: *Detektory zasahující do vozovky, úvod do detekce* [online]. Praha: Dopravní fakulta ČVUT v Praze [cit. 22. 4. 2019] Dostupné z: <https://zolotarev.fd.cvut.cz/mzd/ctrl.php?act=show,file,23843>
- [35] Plzeňský kraj. *Informační systém Plzeňského kraje – Statistiky okresů* [online]. [cit. 23. 4. 2019] Dostupné z: <https://doprava.plzensky-kraj.cz/site/stats>
- [36] Plzeňský kraj. *Informační systém Plzeňského kraje – Co jsou open data?* [online]. [cit. 23. 4. 2019] Dostupné z: <https://doprava.plzensky-kraj.cz/site/page?view=od-about>

Seznam tabulek

Tab. 1: Označení typu detekčního zařízení.....	29
Tab. 2: Označení směru pohybu vozidla	29
Tab. 3: Jednodušší typologie vozidel	30
Tab. 4: Podrobnější typologie vozidel.....	32
Tab. 5: Nová typologie vozidel.....	43
Tab. 6: Rozložení typů vozidel podle typu záznamu; stanice č. 51, ve směru	44
Tab. 7: Rozdělení relevantních dat do skupin podle dělicího znaku.....	46
Tab. 8: Nové rozdělení dat a úpravy formátu	46
Tab. 9: Průměrné rychlosti podle typu zařízení bez filtrování; st. č. 51.....	50
Tab. 10: Ilustrační příklad procentuálního rozložení záznamů podle typu pořízení.....	58

Seznam obrázků

Obr. 1: Vliv velikosti časových bloků na zkreslení podoby dopravy	39
Obr. 2: Několik ilustračních záznamů z původních dat	46
Obr. 3: Ilustrační záznamy v nové podobě	47
Obr. 4: Několik původních záznamů pro ilustraci transformace	48
Obr. 5: Demontrace výsledného formátu časové agregace pro ilustrační záznamy	49
Obr. 6: Demontrace výsledného formátu časové agregace pro všechny záznamy	49
Obr. 7: Demontrace výsledného formátu časové agregace s průměrnou rychlostí	52
Obr. 8: Demontrace výsledného formátu časové agregace s typologií vozidel	53
Obr. 9: Demontrace výsledného formátu časové agregace s typy záznamů	54
Obr. 10: Demontrace výsledného formátu při využití poměrových ukazatelů typů	56
Obr. 11: Demontrace výsledného formátu při přepočtení poměrových ukazatelů typů	56