ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA DOPRAVNÍ

Jakub Krejčí

Analýza požadavků pro tvorbu
výukové databáze dopravních dat

Analysis of Requirements
of Creating Traffic Data Database

Bakalářská práce

2017

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta dopravní
děkan
Konviktská 20, 110 00 Praha 1

K620..... Ústav dopravní telematiky

ZADÁNÍ BAKALÁŘSKÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Jakub Krejčí

Kód studijního programu a studijní obor studenta:

B 3710 – ITS – Inteligentní dopravní systémy

Název tématu (česky): **Analýza požadavků pro tvorbu výukové databáze dopravních dat**

Název tématu (anglicky): Analysis of Requirements of Creating Traffic Data Database

Zásady pro vypracování

Při zpracování bakalářské práce se řiďte osnovou uvedenou v následujících bodech:

- Analýza uživatelských požadavků na výukovou databázi a řešené úlohy.
- Analýza vstupních datových souborů a jejich zdrojů.
- Návrh možných způsobů kontroly datové integrity.
- Návrh základních aplikací dat (zobrazení, přehledy, apod.).
- Návrh přístupového rozhraní uživatelů a jejich rolí.

ČVUT FD

Rozsah grafických prací: dle pokynů vedoucího práce

Rozsah průvodní zprávy: minimálně 35 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)

Seznam odborné literatury: Přibyl, P., Mach, R.: Řídicí systémy silniční dopravy, ČVUT v Praze, 2003, ISB 80-01-02811-9

Přibyl, P., Svítek, M.: Inteligentsní dopravní systémy, BEN 2001, ISBN 80-7300-029-6

Vedoucí bakalářské práce: **Ing. Martin Langr, Ph.D.**

Ing. Ivan Boyarkin

Datum zadání bakalářské práce: **27. října 2016**

(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání bakalářské práce: **28. srpna 2017**

a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia

L. S.

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE Fakulta dopravní -3-

doc. Ing. Pavel Hrubeš, Ph.D.
vedoucí
Ústavu dopravní telematiky

prof. Dr. Ing. Miroslav Svítek, dr. h. c.
děkan fakulty

Potvrzuji převzetí zadání bakalářské práce.

Jakub Krejčí
jméno a podpis studenta

V Praze dne ..... 27. října 2016

Poděkování

Na tomto místě bych rád poděkoval všem, kteří mi poskytli podklady pro vypracování této práce. Zvláště pak děkuji oběma vedoucím mé bakalářské práce, a to Ing. Martinu Langrovi, Ph.D. a Ing. Ivanu Boyarkinovi, za odborné vedení a konzultování mé práce a za rady, které mi poskytovali.

Prohlášení

Předkládám tímto k posouzení a obhajobě bakalářskou práci, zpracovanou za závěr studia na ČVUT v Praze Fakultě dopravní.

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 28. srpna 2017 Jakub Krejčí

……

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta dopravní

Ústav dopravní telematiky

Analýza požadavků pro tvorbu výukové databáze dopravních dat

Bakalářská práce

srpen 2017

Jakub Krejčí

Abstrakt

Předmětem bakalářské práce „Analýza požadavků pro tvorbu výukové databáze dopravních dat“ je shrnutí motivace, cíle a uživatelské požadavky pro tvorbu systému databází určeného k výuce i výzkumu na Ústavu dopravní telematiky a návrh prvotního podkladu pro následující vývoj tohoto nástroje.

Klíčová slova: dopravní data, dopravní databáze, zpracování dat, systémový popis, zdroje dopravních dat

CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Transportation Sciences

Department of Transport Telematics

Analysis of Reguirements of Creating Traffic Data Database

Bachelor thesis

August 2017

Jakub Krejčí

Abstract

The subject of the bachelor thesis „Analysis of Reguirements of Creating Traffic Data Database“ is summary of motivations, visions, goals and user requirements of creating database system which is designed for teaching and research by Department of Transport Telematics and making helpful publication for next development.

Key words: transportation data, traffic data, transportation database, traffic database, data processing, system description, sources of transportation data, sources of traffic data

Obsah

Obsah 4

Seznam zkratek 6

Úvod 7

1 Motivace a uplatnění 9

1.1 Smysl sběru a uchování dopravních dat 10

1.2 Výukový potenciál 10

1.2.1 Předmět 20UITS 11

1.2.2 Další předměty 12

1.2.3 Projektová výuka a závěrečné práce 13

1.3 Výzkumný potenciál 14

2 Uživatelské požadavky 15

2.1 Metodika 15

2.2 Požadavky 16

2.2.1 Požadavky na obsah databáze 16

2.2.2 Požadavky na uspořádání dat 17

2.2.3 Požadavky na přístupový model 18

2.2.4 Požadavky na zobrazení obsahu 19

3 Datové zdroje 20

3.1 Detektory TSK 21

3.1.1 Druhy detektorů 21

3.1.2 Věcný rámec 22

3.1.3 Časový rámec 23

3.1.4 Prostorový rámec 23

3.1.5 Zdroje a vlastnosti dat 24
3.2 Data z plovoucích vozidel – Projekt RODOS 25
3.2.1 Věcný rámec 26
3.2.2 Časový rámec 29
3.2.3 Prostorový rámec 29
3.2.4 Zdroje a vlastnosti dat 29
3.3 Další zdroje 31
3.3.1 Ostatní data popisující provoz 31
3.3.2 Geografická data popisující infrastrukturu 32
3.3.3 Statistická data s vazbou na dopravu 33
3.3.4 Data specifických druhů dopravy 33
3.4 Přístup k zjišťování dat 34
4 Systémový popis databáze dat a metadat 35
4.1 Uživatelé 36
4.2 Data a metadata 37
4.3 Moduly 38
4.3.1 Přístupový modul 38
4.3.2 Ověřovací modul 39
4.3.3 Upload moduly 40
4.3.4 Zobrazovací a Download moduly 42
4.3.5 Uživatelské moduly 44
Závěr 45
Seznam literatury 46
Seznam obrázků 47
Seznam tabulek 47

Seznam zkratek

CEDA	– Central European Data Agency
ČSÚ	– Český statistický úřad
DMM	– Dynamický model mobility
JSDI	– Jednotný systém dopravních informací
NDIC	– Národní dopravní informační centrum
RDS-TMC	– Radio Data System - Traffic Message Channel
	– Rádiový datový systém – kanál dopravních zpráv
RODOS	– Rozvoj dopravních systémů
RSS	– Rich Site Summary
ŘMSD	– Řízení a modelování silniční dopravy
SOAP	– Simple Object Access Protocol
SOKP	– Silniční okruh kolem Prahy
UITS	– Úvod do inteligentních dopravních systémů
TSK	– Technická správa komunikací (Praha)
XML	– Extensible markup language
	– Rozšiřitelný značkovací jazyk
ZABAGED	– Základní báze geografických dat

Úvod

Práce seznámí čtenáře s vizí databázového systému a nadále se podrobněji zabývá prvotní fází tvorby databáze dopravních dat. Abychom porozuměli problematice dopravních dat, nutnosti jejich sběru, skladování a zpracování, musíme nahlédnout do historie. Provoz dopravy, jak je známo, v důsledku vývoje technologií s plynoucím časem rychle narůstá.

S nárůstem automobilové dopravy ve dvacátém století začala růst i vytíženost silniční infrastruktury. Pro její lepší správu si dopravní inženýrství vyžaduje znalost této vytíženosti, jejího nárůstu i její příčiny, čemuž pomáhá sběr dopravních dat.

Historie automatizace sběru dopravních dat sahá do roku 1937, kdy byl představen jeden z prvních automatických čítačů dopravy. Každou hodinu vytiskl na proužek papíru intenzitu projetých vozidel. S postupem času se sběr dat vyvíjel. Dnešní dopravní detektory jsou napojeny na telematické systémy a posílají data do dopravních center, kde jsou automaticky zpracovány. Výsledkem jsou automatická vyhodnocení dopravní situace a varování před nežádoucími situacemi, například vzniku kongescí. Tento rozvoj nenastal pouze v rámci silniční dopravy, ani v rámci dopravy všeobecně, ale obecně ve sběru dat. V současnosti je v souvislosti s daty často zmiňovaný pojem „Big data“. Ten popisuje jakákoliv data, která narůstají exponenciálně, vyžadují rychlé zpracování, jsou různorodá a jejich struktura nemusí být zřejmá. Příslušnost dopravních dat k pojmu „Big data“ je nevyvratitelná. Pokud budeme uvažovat nekomprimovaná dopravní data v České republice, přibyde jich denně v řádu gigabajtů. Ty je potřeba chytře a správně analyzovat. Odměnou za tuto analýzu jsou nové inovace na poli dopravních systémů.

Mezi další potřeby, které vedly ke zpracování této práce, patří potřeba zjednodušení procesů zpracování dat. To si lze představit jako unifikovaný proces podobný pro většinu dopravních dat.

Dnešní nástroje na zpracování dat (včetně těch méně profesionálních) umožňují připojení přímo k databázi, což významně ulehčuje práci s daty. Uživatel, který data zpracovává bez databázového systému a těchto nástrojů bojuje s různorodostí zdrojů a dat. Hledáním

nesrovnalostí v datech a jejich následnou opravou stráví uživatel až 75 % času celkového zpracování dat.

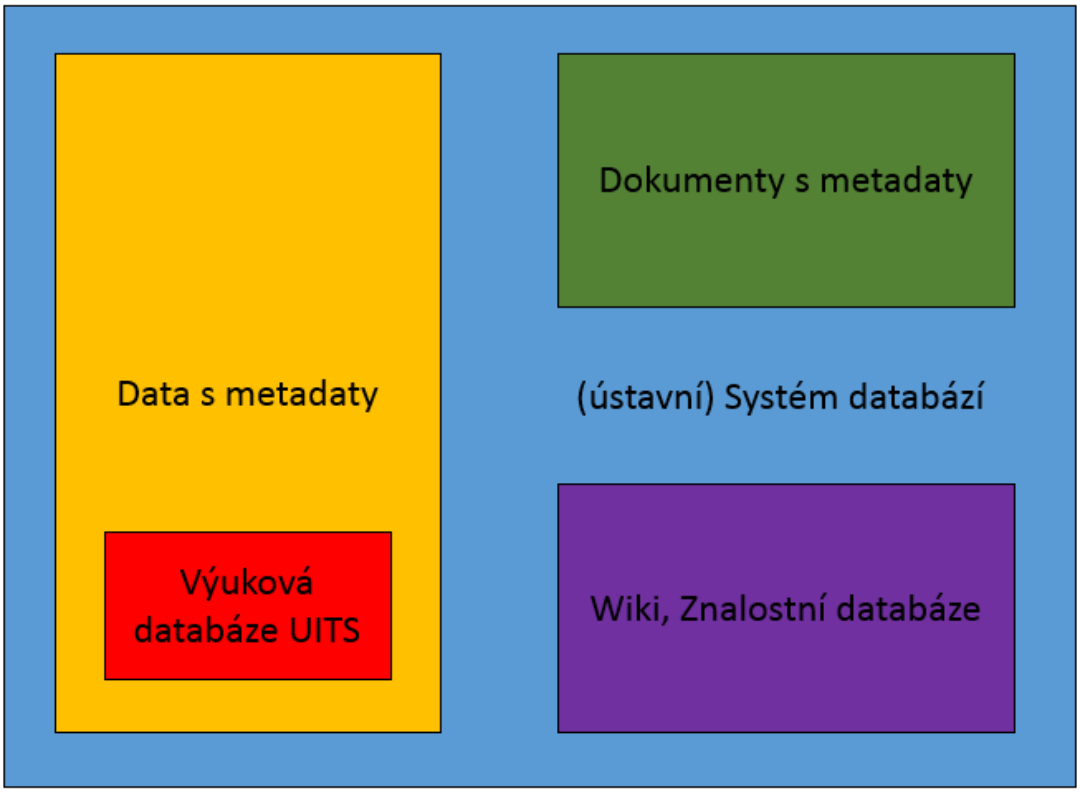
Vznikem systému databází též vznikne jasně vymezený prostor pro výsledky výzkumné práce.

Ústav dopravní telematiky Fakulty Dopravní ČVUT v Praze se problematikou dopravních dat zabývá, a má několik zdrojů dat k dispozici. Vznikla zde potřeba univerzálního zpracování dat což vyvolalo vznik potřeby vytvoření vhodného nástroje.

Cílem této práce je shrnout motivace, vize, cíle a základní uživatelské požadavky. Poté lze vytvořit prvotní náhled na vytvářený nástroj a shrnout tyto informace do jedné publikace, která bude sloužit jako podklad pro další vývoj tohoto nástroje. Tato práce má netypický charakter spočívající především v netradičních zdrojích informací. Autor zde využívá především technických příruček a rozhovorů s uživateli.

1 Motivace a uplatnění

Během poznávání problematiky této práce se čtenář setká s několika pojmy databáze. Je velmi důležité tyto pojmy vymezit. Základní systém databází se skládá ze tří typů databází. Data reprezentovaná hodnotami, které jsou skladované v tabulkách, data obsažená v dokumentech a wiki web obsahující znalosti. Bude-li řeč o výukové databázi UITS, bude se jednat převážně o čistě klasický datový druh databáze. To je přehledně ukázáno na následujícím obrázku.



Obrázek 1- Schéma systému databází;
zdroj: archiv autora

Prvním krokem je přiblížit čtenářovi smysl ústavního systému databází, rozšířit povědomí o uplatnění a shrnout motivační myšlenky. Tato kapitola se zabývá důvody, proč má smysl shromažďovat data a znalosti. Dále se kapitola zaměřuje na problematiku výuky

a výzkumu, respektive na inovace, které v těchto odvětví mohou nastat díky tomuto systému databází.

1.1 Smysl sběru a uchování dopravních dat

Nejprve je nutno se zabývat důvody, proč se shromažďují dopravní data obecně. Kapitola vychází z publikace Traffic Engineering. [1]

Pro dopravní inženýrství je důležité znát stav dopravy na komunikacích, a to ať už historický, aktuální i předpokládaný v budoucnosti. Tyto stavy se popisují dopravními veličinami, které jsou snímány nejen dopravními detektory, ale i flotilovými systémy či mobilními telefony běžných uživatelů dopravního systému. Tato dopravní data však obsahují leckdy mnoho zbytečných údajů. Ty je potřeba správně sbírat a s ohledem na využití také z důvodu velkých objemů dat správně agregovat.

Je důležité sbírat dopravní data pro určení trendů a následné určení dopravního stavu v budoucnosti. Zrovna tak data dopravní nehodovosti mohou po analýze určit problémové lokality, naznačit směr jejich řešení a při znalosti trendu nehodovosti lze problém řešit ještě předtím, než vznikne.

Znalost přepravních vztahů obyvatel i zboží, potřeby parkování a tím tedy celkové poptávky může vést k co nejefektivnějšímu dopravnímu systému např. města. S tím souvisí i určování a predikce vlivů dopravy na okolí. Podobně se dají porovnat i zásadnější změny v dopravní infrastruktuře a vyhodnotit vlivy těchto změn.

Na mnohem nižší úrovni sběru dat se dají kalibrovat různé dopravní parametry – například v řadiči křižovatky mohou data z dopravních detektorů (třeba časová mezera) vést k zefektivnění řízení křižovatky.

1.2 Výukový potenciál

Po shrnutí obecných motivací shromažďování dat se práce dále bude zabývat specifičtějšími motivacemi, konkrétně perspektivou zamýšleného systému databází s ohledem na výukovou činnost Ústavu dopravní telematiky

Pro výuku je typické předávání informací směrem od vyučujících ke studentům. Kromě informací předaných v lekcích se předávají také související dokumenty a zpravidla i souhrn v podobě prezentací z průběžných lekcí. V současnosti, má každý předmět svá pravidla zveřejňování učiva. Existují sice předměty, které mají své webové stránky, ale na druhou stranu existují také předměty, kde je souhrn informací a přednášek z lekcí rozposílán e-mailem téměř před zkouškou. V případě nahrazování absence si student musí vyžádat prezentaci u vyučujícího, což může být ještě navíc zkomplikované, pokud předmět vyučuje více vyučujících. Student většinou ani nemůže detailněji nahlédnout na látku nastávající lekce.

Znalostní databáze ve smyslu wiki webu rozšiřují možnosti studentů tyto informace přijímat. Jediný společný prostor přístupný pro studenty, který by obsahoval podstatné informace a dokumenty či například prezentace lekcí, by zcela jistě zlepšil kvalitu vzdělávání. Zlepšení spočívá především v přístupu k informacím a v jednotvárnosti tohoto systému společného pro více předmětů. Tento společný koncept není pouze sjednocením složek předmětů z fakultního disku, ale je systematizovaný a unitární.

V následující kapitole je uveden příklad plánovaného využití databázového systému v jednom konkrétním předmětu.

1.2.1 Předmět 20UITS

V zimním semestru roku 2016 v rámci nové akreditace předmětu Úvod do Inteligentních dopravních systémů (dřívější kód 20UIS) byla část semestrální práce věnována zpracováním dat ze strategických dopravních detektorů. V úloze měli studenti zpracovat týden dat z detektorů CollectR na území Prahy v 18 lokalitách. Pětiminutová data z těchto detektorů jako rychlost, intenzita a obsazenost (podrobnější specifikace dat se nachází v kapitole 3.2) studenti stáhli z předem připravených dokumentů obsahující data v jazyce XML, dále tyto data pospojovali, nahráli do tabulkového editoru, kde data filtrovali, opravovali chyby dat a dále zpracovávali tyto data, čímž tvořili technickou zprávu.

Hlavní přínos ústavního systému databází spočívá v přístupu studentů k datům. Ten by měla zajišťovat výuková databáze UITS. Bez ní se musí na ústavu manuálně stahovat dopravní data a ty dále distribuovat studentům přes fakultní disk, či pomocí internetu.

Student je nucen překopírovat data v podobě sedmi textových souborů z fakultního disku na svůj a úplně tím ztrácí ponětí o databázovém přístupu k datům. Datová databáze, která zálohuje a uchovává data přímo od jeho poskytovatele, je může se stejnou jednoduchostí poskytovat studentům. S pomocí výukové databáze si studenti stáhnou přímo ta data, která potřebují a díky logování přístupu se tomu nemůžou vyhnout, neboť bude dohled nad tím, že si každý student skutečně stáhne svá data. Student tak taky bude přistupovat do databáze a bude mít přehled, jaká data jsou k dispozici. Import dat zajistí například přes rozšířené funkce tabulkových editorů, které se napojí přímo na databázi. Data se tedy mohou vybírat a porovnávat bez vzniku zbytečných duplicitních souborů.

Samotný popis sítě detektorů a detektorů samotných je další ukázkový příklad případného obsahu wiki webu. Student by zde mohl najít detailní popisy detektoru, měřených veličin a podoby dat, které poskytuje. Informovanost studentů o dostupnosti konkrétních dat rozvíjí jejich tvůrčí a analytické schopnosti, protože jim dává možnost data porovnávat či jinak analyzovat. Budeme-li uvažovat naplnění databáze z mnoha datových zdrojů, může si student na základě wiki webu vybrat vhodné zdroje dat k porovnání. Vzhledem k faktu, že množství veškerých dat i množství měřených druhů dat s plynutím času roste, a vzhledem k možnosti získávání znalostí o datech současně, je tato úloha pro studenta velmi přínosná.

1.2.2 Další předměty

Uplatnění tohoto systému databází by se mohlo najít i v jiných předmětech. Záleží však na vedení předmětu, zda by tento koncept využilo. Prozatím neproběhlo další zjišťování ve větší míře. To je možné provést až po zpracování alespoň jednoho konkrétního řešení jako příkladu. Mezi předměty [2], které jsou tematicky blízko dopravním datům, patří následující:

Z hlediska procvičování databázového zpracování dat:

- Algoritmizace a datové struktury	14AS	Koukol
- Analýza a zpracování dat	14AZ	Svoboda
- Databázové systémy	14DB	Kaliková
- Informační systémy v dopravě	14ISYD	Kalika

Z hlediska využití dopravních dat:

- Řízení dopravního uzlu a linie	20RU	Tichý
- Matematické metody	11MAMY	Přikryl
- Řízení silniční dopravy	20RISI	Tichý
- Aplikovaná telematika	20APLT	Svítek
- Teorie řízení a spolehlivosti v dopravě	20TRAS	Faltus
- Geografické, informační, lokalizační a navigační systémy	20GIL	Hrubeš

- a další …

1.2.3 Projektová výuka a závěrečné práce

V současnosti probíhá projektově zaměřená výuka na Ústavu dopravní telematiky v 11 projektech, kde mezi největší projekty patří projekt Řízení a modelování silniční dopravy (jedná se i o projekt autora BP, proto se jím bude autor více zabývat).

Obecně se od povinné účasti na projektu očekává vlastní práce v daném tématu, ať už jde o pomoc s výzkumem ostatních účastníků projektu či výzkum vlastní. Tento studentský výzkum může být stejně přínosný jako každý jiný výzkum, avšak bohužel není nijak masivně sdílen.

V projektu ŘMSD se každý student musí pro splnění předmětu zakončeného zápočtem věnovat nějaké odborné činnosti, která pak může být prezentována na společných pravidelných schůzkách, avšak většinou pouze ve skupině účastníků tohoto projektu.

Znalostní databáze, kam by každý student nahrál svůj výstup, by přinesla širší povědomí o činnosti probíhající v projektu a informace by byly jednoduše použitelné i pro ostatní uživatele, kteří nejsou v přímém kontaktu s účastníky projektu.

Projektová výuka je též doporučená a spojovaná s tématem závěrečné práce. Při psaní této práce by si studenti mohli jednoduše dohledat potřebné informace i data. Pro lepší názornost bude uveden příklad. Běžný typ bakalářské práce na ústavu je návrh signálního

plánu křižovatky. Student by si v tomto případě nalezl danou křižovatku na wiki webu, kde by mohl být popis dostupných dat z detektorů. K těmto datům by mohly přímo přistupovat i simulační programy. Hlavní předností by však bylo neomezené vyhledávání dalších potenciálně použitelných dat na jednom místě.

1.3 Výzkumný potenciál

Vedle výuky se též najde perspektiva zamýšleného systému databází i ve výzkumné činnosti Ústavu dopravní telematiky.

Pro výzkum v dopravní oblasti je typické zpracování a práce nejen s daty, ale například i s dokumenty.

Skladování veškerých dokumentů a dat, které nejsou k dispozici kdykoliv, přinese větší informovanost všech uživatelů systému databází a poskytnou pohodlným způsobem data i znalosti potenciálně potřebné k výzkumným činnostem.

Největší přínos je však ukryt v pohledu na všechna data dohromady. Ty mohou tvořit dynamický model mobility. Jedná se o matematický model, který má široké spektrum vstupních dat. Mimo výpovědi o aktuální či predikované dopravní situaci by měl zároveň vypovídat i o dalších oblastech spojených s dopravou jako například o energetické spotřebě, emisní zátěži, přepravě zboží, pohybu chodců a podobně. Ten může vzniknout postupným zkoumáním a zapojováním jednotlivých datových zdrojů.

2 Uživatelské požadavky

Po shrnutí veškerých idejí a vizí je na čase změnit měřítko pohledu na navrhovaný systém a podívat se na něj o něco podrobněji. Tato kapitola je souhrnem uživatelských požadavků od případných uživatelů a osob se zájmem na vývoji tohoto systému databází. Požadavky slouží k pochopení databázového systému a jsou nutné pro jeho správný návrh, proto musí být všechny řečeny před začátkem procesu navrhování.

2.1 Metodika

Pro zjišťování uživatelských požadavků zvolil autor oslovení zamýšlených budoucích uživatelů systému databází. Mezi ně patří pracovníci Ústavu Dopravní telematiky, u nichž se očekává případné využití tohoto systému.

Příprava na první schůzku spočívala v sepsání několika zásadních témat:

- využitelnost (vymyslet všechny možné uplatnění)
- rozsah datový (jaké datové zdroje použít)
- rozsah geografický (v rámci jakého území – ČR/Praha?)
- rozsah časový (objem vůči užitečné době)
- forma a přístup (použití web/aplikace, heslo)
- datové formáty (csv či xml)
- přehledy a zobrazení (funkčnost detektorů, dopravní situace)
- další (řešení opravování dat)

První schůzka proběhla ve středu 21. června 2017 v 9:00 s účastníky:

- Ing. Martin Langr Ph. D.
- Doc. Ing. Pavel Hrubeš Ph. D.
- Ing. Jiří Růžička
- Ing. Přemysl Derbek Ph. D.

V úvodu schůzky bylo ujasněno, že původní představa náplně této Bakalářské práce byla na moc podrobné úrovni a ubírala se příliš technickým směrem. Počítalo se se zapojením

již dostupných zdrojů dat a nehledělo se tolik na motivace a vize. Původní otázky a cíle se týkaly spíše výukové databáze nežli celého systému.

Bylo zde také doporučeno, aby se práce dále na zadané téma soustředila v obecnější rovině. Technické detaily budou konkrétní ke konkrétním datovým zdrojům při jejich následném zkoumání.

Schůzka probíhala přibližně 2 hodiny a měla podobu brainstormingu, jenž by se jednoduše dal rozdělit do čtyř nově vzniklých tematických částí:

- Obsah
- Uspořádání dat
- Přístup
- Zobrazení obsahu

Tyto nové tematické celky obsahují podstatné části původně připravených bodů. Z náplně brainstormingu jsou vynechány technické detaily konkrétních datových zdrojů, které jsou pak analyzovány zvlášť v následující podkapitole.

Dále probíhaly individuální schůzky s vedoucím práce Ing. Martinem Langrem Ph. D. a Doc. Ing. Pavlem Hrubešem Ph. D.

2.2 Požadavky

2.2.1 Požadavky na obsah databáze

Co se požadovaného obsahu týče, lze jej rozdělit na složku dat (tedy měřených veličin či statistických údajů) a složku metadat, která slouží jako popis daných dat pro uživatele, například popisu dosavadního zpracování, uvedení zdroje dat a podobně.

Data

V databázi je třeba počítat se širokým spektrem druhů dat. Má-li dobře sloužit, musí brát zřetel na široké souvislosti mezi jednotlivými typy dat a umět je poskytnout uživateli. Jedná se o data všech druhů dopravy i další data s dopravou zdánlivě nesouvisející, jako

například data demografická a statistická. Výzkumu bližších informací o datech se věnuje kapitola 3.

V případě, že přibude nový zdroj dat, musí jít do databáze pohodlně vložit, a to i v případě, že se jedná o zcela nový typ dat, který se nepodobá žádnému z použitých.

Některá data, zejména ta dostupná online, nemusí být pořád stahována do databáze, nýbrž postačí, budou-li skladovány pouze data za poslední období. Toto určité období se musí posoudit individuálně ke každému konkrétnímu typu dat. Toto množství se určí s ohledem na velikost dat a jejich potřebu je skladovat. Použití dat pro výukové potřeby si vyžádá výpočet závislý na počtu studentů a zadání úlohy. Předpokladem pro omezování doby skladování je neomezený online přístup k libovolně starým datům.

Je zájem o veškerá související data z území České Republiky, i pokud se jedná o data týkající se pouze jednoho menšího území, například města.

Metadata

Skladování metadat je neméně důležitou součástí databáze. Spolu s daty musí být zachovány informace s nimi související. V případě, že se jedná o online zdroj dat, musí být dostupné informace, o jaká data vlastně jde. Mezi hlavní metadata patří popis tří měřítek – věcného, časového a prostorového rámce. Dále je potřeba vést informace o dostupnosti zdroje a jiných technických detailů. V případě dokumentů, to může být informace o autorovi, či zdali je dokument dostupný v tištěné podobě a kde jej lze nalézt.

2.2.2 Požadavky na uspořádání dat

Data získaná přímo od poskytovatele se musí uchovávat odděleně od dat upravených uživatelem, i přes to, že jsou nezpracovaná. Jejich zpracování nemusí být pro všechny použití správné, a proto musí být vždy umožněn přístup k datům původním.

Každá datová sada bude mít tedy pevně dané (nesmazatelné, neupravitelné) verze, mezi které budou patřit originální datové sady a jejich zpracované verze, které jsou však dlouhodobě ověřené. Vedle pevných verzí datových sad musí být i verze dočasné, tedy upravené verze (zpracovaná data) která ještě nejsou ověřena. K těmto neověřeným datům není doporučen veřejný přístup a měl by o něm rozhodovat autor zpracování.

Ověřování těchto zpracování (či modelů) by mělo probíhat po vícenásobném odborném posouzení a porovnáváním s ostatními konkurenčními modely.

Při ověřování modelu je třeba vedle posouzení z hlediska správnosti modelu, posoudit model i z hlediska pravdivosti, což bývá často opomíjeno. Máme-li tedy model, který opravuje a dopočítává chybějící data (například při výpadcích spojení detektory) musí být výstupem modelu nejen data správná a zdánlivě smysluplná, ale i pravdivá a co nejvíce totožná s realitou.

Ověřování dat je potřeba řešit měnitelným aplikačním modulem z důvodu možné změny způsobu opravování.

Přestože je oprava surových dat v průběhu veškerého zpracování dat pro člověka časově nejnáročnější, je zde velice snadné udělat velice zásadní chybu, která by měla vliv na výstup. Z tohoto důvodu vzniká potřeba věnovat velkou pozornost případným opravovacím modulům.

Mělo by být počítáno i s možností plnit databázi dokumentů aktualizačními technikami jako je například RSS. To by pomohlo plnit databázi pravidelnými dokumenty jako například ročenkami institucí, které obsahují užitečná data i informace.

2.2.3 Požadavky na přístupový model

Přístupový model by měl mít vymezitelný seznam uživatelů a jejich práv na konkrétní části databáze ať už se jedná o zápis, čtení, editaci či mazání.

Uživatelé by neměli data zbytečně stahovat, ale pouze prohlížet na serveru.

Měl by být veden přehled o přístupech uživatele do databáze a činnostech, které byly prováděny (log). Ten je důležitý jednak pro řešení následků případných neoprávněných činností. Pro použití v předmětu UITS by pak sloužil pro kontrolu toho, zda si každý student stáhl svá data sám.

2.2.4 Požadavky na zobrazení obsahu

Je vhodné počítat se skutečností, že pro veškerá data je potřeba navrhnout nějaké čitelnější zobrazení. Z hlediska uživatelů pracujících na výzkumu by bylo vhodné zobrazovat veškerá data. To se však netýká komplexnějších zobrazení vytvářených na míru pro případného „zákazníka“.

V případě dat by bylo vhodné zobrazovat základní charakteristiky, jako je počet řádek či počet unikátních hodnot.

3 Datové zdroje

Nesmírně důležité při úvahách o tvorbě datové databáze je věnovat kapitolu datům na vstupu do databáze a jejich zdrojům. Nejprve budou popsána data v obecné rovině, včetně základního dělení. Poté bude věnován prostor dvěma zdrojům, které už byly podrobeny prvotní analýze autorem této práce. Po seznámení s prvními zdroji dat bude naznačena šíře dalších potenciálně vhodných zdrojů.

Orientaci v množství potenciálně vhodných dat a zdrojů usnadní tabulka s přehledem těchto dat. V budoucí přehledové tabulce by byly uvedeny zdroje dat i s přehledem jejich charakteristických vlastností, coby společných atributů. Tímto se zaručí komfortní vyhledávání v případě potřeby zobrazení něčím charakteristických dat, například bude-li potřeba vědět, které všechny zdroje obsahují data týkající se železniční dopravy, či se budou srovnávat data z plošných území např. krajů.

Kromě atributů, které by mohly vzniknout na základě následující myšlenkové mapy, je zapotřebí i detailní ucelený popis podobný následujícím kapitolám.

Obrázek 2 - Myšlenková mapa jako náznak rozdělení zdrojů;
zdroj: archiv auora

Vedle zdrojů, které jsou již známy úplně by v tabulce měly být zdroje známé i částečně, i kdyby jen jejich název a popis. To by zaručilo, že pozdějším výzkumem by se množství dat mohlo zvětšovat.

3.1 Detektory TSK

Dopravní detektory ve vlastnictví Technické správy komunikací Praha se dělí na několik druhů. Jedná se o křižovatkové i strategické detektory na celém území Prahy. Technickým zprostředkovatelem těchto dat je firma ELTODO. S touto firmou autor spolupracoval, za účelem zjištění podrobností o těchto datech, nicméně narazil na neinformovanost a zmatek. Nejsou ani známy podrobnosti o většině detektorů, jejich polohy ani informace o současném využití těchto dat.

Autor osobně prováděl analýzu těchto dat v rámci příprav praktické úlohy předmětu 20UITS v srpnu 2016. Z podkladů tohoto zkoumání [3] následně čerpal i v této práci.

3.1.1 Druhy detektorů

Jednotlivé druhy detektorů mají různé vlastnosti, měřené veličiny a podobně. Základní přehled vyjadřuje následující tabulka. Podrobněji popsaná jednotlivá měřítka budou popsána v následujících kapitolách 3.1.x.

	Věcný rámec								Prostorový rámec		Časový rámec	
	souhrnné veličiny				klasifikované veličiny				detektorů (jízdních pruhů)	lokalit	interval poskytování	
Druh detektoru:	počet	obsazenost	rychlost	hmotnost	počet	obsazenost	rychlost	hmotnost			surové	agregace
CollectR	ANO	ANO	ANO	NE	1-2	NE	1-2	NE	348	cca 100	5 min	1 hod
MP	ANO	ANO	ANO	NE	0-1-2-3	NE	NE	NE	187	86	5 min	1 hod
MUR	ANO	ANO	ANO	NE	0-1-2-3	NE	NE	NE	63	22	5 min	1 hod
SCALA	ANO	ANO	ANO	NE	NE	NE	NE	NE	4565	259	1,5 min	1 hod
SDDU	ANO	ANO	ANO	NE	0-1-2-3	NE	NE	NE	47	22	5 min	1 hod
Traficam	ANO	ANO	NE	NE	NE	NE	NE	NE	78	cca 20	5 min	1 hod
WIM	ANO	ANO	ANO	ANO	1 AŽ 13	1 AŽ 13	1 AŽ 13	1 AŽ 13	13	7	5 min	1 hod

Tabulka 1- Přehled detektorů TSK; zdroj: archiv autora

Detektory CollectR a Traficam jsou od výrobce FLIR, zatímco ostatní (MP, MUR, SDDU, WIM) jsou od české společnosti Camea. Křižovatkové detektory pak spravuje provozovatel dané křižovatky (Eltodo, AŽD, …).

3.1.2 Věcný rámec

Jednotlivé detektory měří a posílají údaje až o čtyřech veličinách. Patří mezi ně intenzita, obsazenost detektoru, rychlost a hmotnost. Hodnoty jsou vedeny jak pro různé kategorie vozidel zvlášť, tak i souhrnně pro všechny, tedy vážené průměry pro rychlosti a součty pro intenzity, obsazenosti a hmotnosti. Přiřazení veličin, druhů detektorů a případné klasifikace lze vyčíst z předchozí tabulky.

Rozdělení kategorií vozidel

Různé typy detektorů, měří kromě souhrnné hodnoty všech vozidel i hodnoty pro specifické kategorie vozidel. Tyto způsoby se liší u různých druhů detektorů (viz předchozí tabulka)

- Rozdělení 0-1-2-3 se vyskytuje u typů MP, MUR a SDDU. Jednotlivé kategorie dle manuálu firmy Eltodo:
    - třída 0 - Nezařazená vozidla
    - třída 1 - Osobní vozidla
    - třída 2 - Lehká nákladní vozidla
    - třída 3 - Těžká nákladní vozidla
- Rozdělení 1-2 se vyskytuje u typu CollectR. Jednotlivé kategorie dle manuálu firmy Eltodo:
    - třída 1 - Osobní vozidla
    - třída 2 - Nákladní vozidla
- Rozdělení 1 až 13 se vyskytuje u typu WIM, jde o rozdělení podle počtu náprav, jejich rozvorech a celkové délce vozidla. Podrobnější informace nebylo možné zjistit.

3.1.3 Časový rámec

Každý druh detektoru má dle tabulky časový interval, za který posílá informace. Jedná se o 90 vteřin u křižovatkových detektorů napojených na ústřednu SCALA a pětiminutový interval u ostatních – strategických detektorů (umístěných v mezikřižovatkových úsecích sběrných komunikací). Tyto intervaly jsou i v rámci jednoho druhu posunuty a začínají spontánně (např. jeden detektor udává informaci pro čas 0:02:12 – 0:07:12 a další detektory jinak).

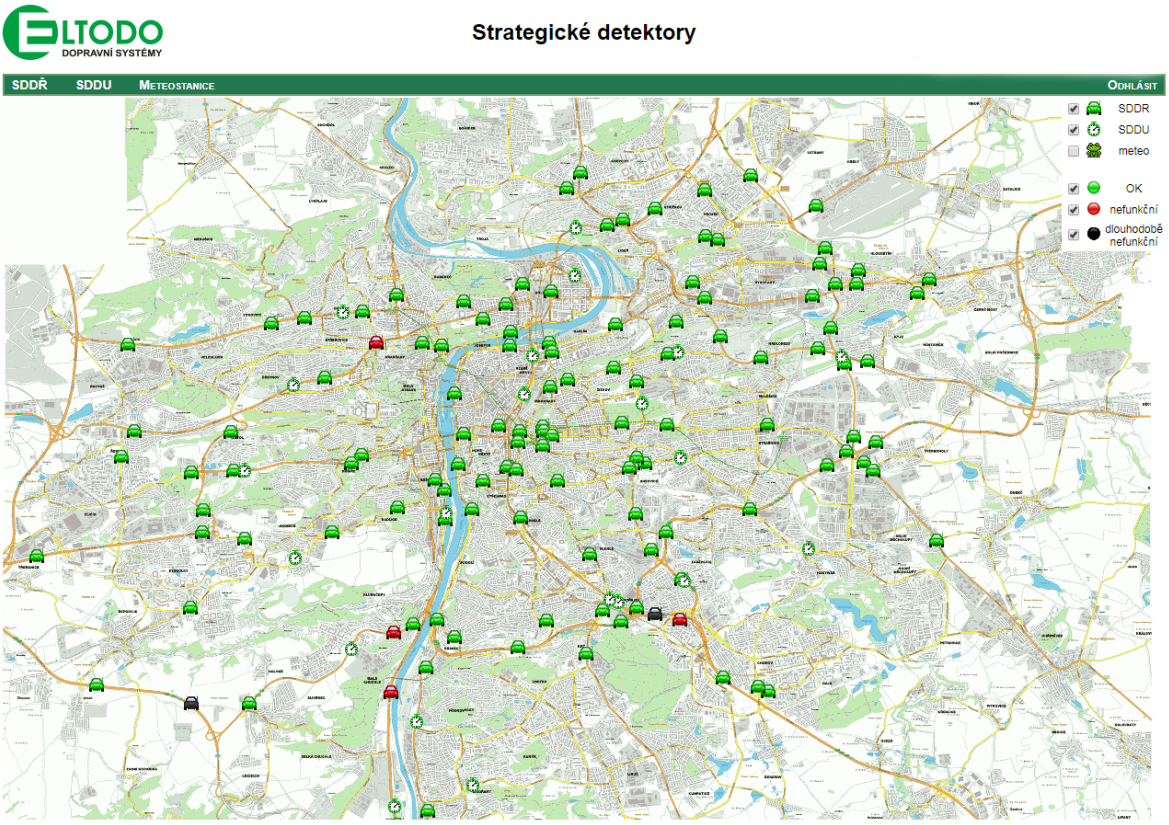
Surová data detektorů výrobce Camea umožňují zpětnou dostupnost dat pouze 1 hodinu zpět, zatímco detektory CollectR umožňují zpětnou dostupnost 8 dní.

Dále se dle druhu liší i datový formát, respektive konstanta pro časový posun u detektorů CollectR je „Z“ namísto obvyklých „+02:00“.

3.1.4 Prostorový rámec

Jelikož se jedná o detektory TSK Praha, dá se předpokládat jejich umístění v Praze. Detektory se dají rozdělit podle prostoru, který popisují – na liniové (SDDU, MUR) a bodové (ostatní). Bodové detektory se dále dělí na křižovatkové (indukční smyčky ve vjezdových jízdních pruzích světelných křižovatek) a strategické (umístěné v mezikřižovatkových úsecích sběrných komunikací).

Přesné lokace jsou zatím známy jen u detektorů Traficam, CollectR a SDDU díky mapovému rozhraní [4] firmy Eltodo. Přibližné rozložení těchto detektorů je znázorněno na následující mapě.



Obrázek 3 - Mapa strategických detektorů Eltodo
zdroj: stratdet.eltodo.cz

3.1.5 Zdroje a vlastnosti dat

Data jsou poskytována přes protokol Transmitter 3.5, který je popsán v manuálu firmy Eltodo [5]. Poskytovaná data jsou v jazyce XML. Nejvyšším elementem je zpráva „msg“ s atributem typu – data. Tato zpráva obsahuje data od několika detektorů – to závisí na poskytujícím systému. Tyto zprávy jsou posílány za sebou v čase, dle aktuálního zpracování zpráv z detektorů v poskytujícím systému.

Vzorek dat

Jedna zpráva <msg> může obsahovat i více zpráv od detektorů <detector>. Element <detector> s atributy ID detektoru a druh detektoru obsahuje stavy veličin za jedno dané časové okno daného detektoru.

Vzorek vypadá následovně:


```
<msg type="data">
<detector id="R016010-V1" type="CollectR">
<datetime format="yyyy-MM-dd HH:mm:ssK">2016-07-28 10:27:29+02:00</datetime>
<start format="yyyy-MM-dd HH:mm:ssK">2016-07-27 10:54:48+02:00</start>
<stop format="yyyy-MM-dd HH:mm:ssK">2016-07-27 10:59:48+02:00</stop>
<count>
        <all>87</all>
        <class id="1">56</class>
        <class id="2">31</class>
</count>
<speed>
        <all>40</all>
        <class id="1">40</class>
        <class id="2">40</class>
</speed>
<occupancy>
        <all>24</all>
</occupancy>
</detector>
</msg>
```


Takovýto text v kódování ANSI má u detektorů CollectR velikost cca 530 kB na 1 detektor (4 jízdní pruhy) a den. Všechny detektory CollectR v Praze tedy potřebují zhruba 45 MB/den. U všech detektorů je objem dat větší, zhruba 2 GB/den pro všechny detektory. Příčinou budou nejspíše křižovatkové detektory, kterých je dle tabulky drtivá většina a na rozdíl od ostatních posílají data po 90 sekundách namísto 5 minut.

3.2 Data z plovoucích vozidel – Projekt RODOS

Data z plovoucích vozidel nesou údaje o stavu silniční sítě v ČR. Patří mezi ně rychlost, cestovní čas, úroveň dopravní zátěže a zpoždění. Ty jsou vypočítány z odeslaných údajů o poloze z vozidel, které jsou vybaveny flotilovými systémy. Tyto údaje se dále zpracovávají a vytváří se dopravní model. Bližší podrobnosti o tomto zpracování však nejsou známy, nejspíše z důvodu ochrany know-how.

Zpracovatelem těchto dat je centrum RODOS, složené z odborníků z akademické sféry, předních českých soukromých firem i státní správy a komplexně řeší problematiku

modelování, řízení a optimalizace dopravy budoucnosti. Jádrem centra RODOS, který v následujících 6 letech řeší 6 základních a vzájemně se doplňujících výzkumných tematických okruhů z oblasti dopravy, je Dynamický Model Mobility ČR (DMM integruje dynamický model pohybu osob, vozidel, zboží a s tím souvisejících informací v rámci celého území ČR). Data z plovoucích vozidel jsou využívány jako jeden z pilířů DMM.

Autor osobně prováděl analýzu těchto dat a porovnával je s daty z detektorů na SOKP Praha v rámci studie pro centrum RODOS v únoru 2017. Ze vzniklého dokumentu [6] následně čerpal i v této práci.

3.2.1 Věcný rámec

Nejzákladnějším dělením posílaných hodnot je rozdělení do tří stavů:

- stav v danou dobu aktuální
- stav v danou dobu obvyklý
- stav při volném proudu

V každém stavu se pak jedná o hodnoty veličin – rychlosti, cestovní doby, úrovně dopravní zátěže a zpoždění. Po analýze dat bylo zjištěno, že jediná nezávislá veličina je cestovní doba, poněvadž rychlost je dopočítaná právě z cestovní doby a konstantní délky daného úseku. Totéž platí pro zpoždění coby rozdílu aktuální cestovní doby a cestovní doby při rychlosti volného proudu.

Stav volného proudu

Ve stavu volného proudu se setkáváme s pojmem „Free Flow rychlost“. Rychlost Free Flow, je definována jako rychlost při nízké intenzitě, kdy se vozidla navzájem neovlivňují. Hodnota je stanovena jako průměrná rychlost 20 % nejrychlejších vozidel v segmentu (bez odlehlých hodnot). Z hlediska vyhodnocených dat je to zároveň rychlost limitní – maximální. V datech z plovoucích vozidel se pro daný segment nevyskytuje žádná vyšší hodnota. Hodnota rychlosti volného proudu většinou nezávisí na denní době a v tomto stavu jsou hodnoty veličin konstantní v kterémkoliv časovém úseku, tedy například během dne je odesláno 1440 minutových hodnot rychlosti, cestovní doby a zpoždění, které jsou po celý den konstantní. Toto však nemusí být pravidlo, pokud by se hodnoty rychlosti při

volném proudu odvíjely od denní doby, například v závislosti na světelných a klimatických podmínkách.

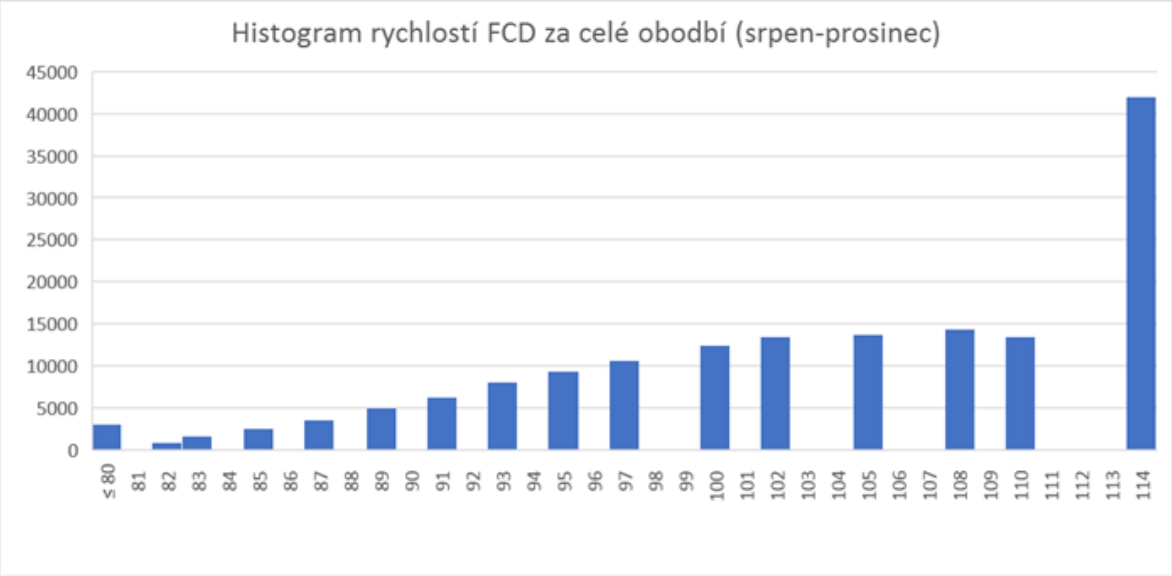
Normální stav

V obvyklém stavu se setkáváme s hodnotami, které jsou průměrné a typické pro daný segment v danou dobu. Zpracování probíhá vhodným způsobem, který odstraňuje šumy v datech a umožňuje data správně vyhodnotit. Tato informace vychází z příručky pro data služby RODOS. [7] Podrobnější informace jsou nedostupné.

Aktuální stav

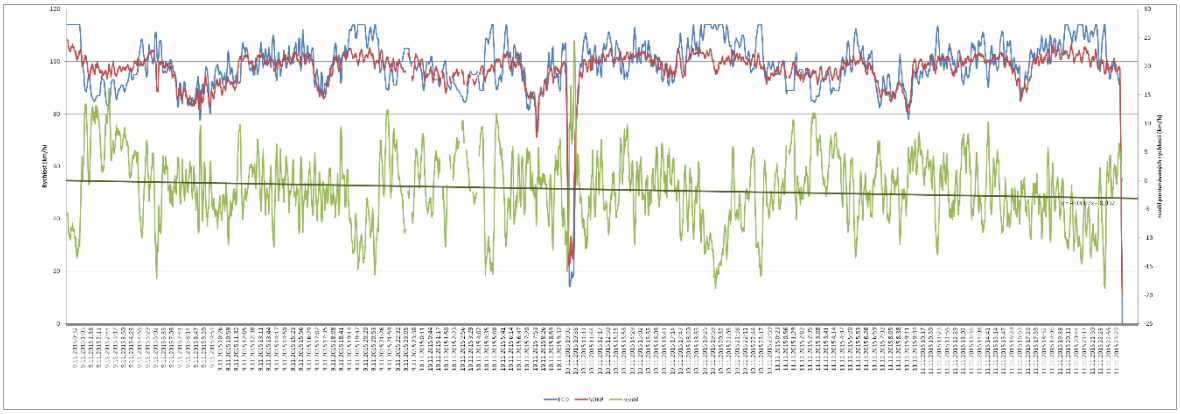
V aktuálním stavu jsou sdělovány hodnoty veličin pro daný časový úsek. Pokud je skutečná aktuální rychlost vyšší než vyhodnocená rychlost Free Flow, nahrazuje ji právě rychlost Free Flow, neboť je to teoreticky nejvyšší možná rychlost v daném úseku. Je zde také navíc úroveň dopravní zátěže, která se počítá na základě podílu aktuální průměrné rychlosti a rychlosti volného průjezdu. [7]

Hodnota aktuální rychlosti je generovaná v krocích po 2-3 km/h, čehož si lze všimnout z dlouhodobého histogramu na jednom úseku. Specifikum těchto dat je též zaokrouhlování rychlostí vyšších, než je rychlost volného proudu – to lze v grafu vidět v podobě vyšší četnosti rychlosti volného proudu.



Obrázek 4- Histogram rychlostí FCD
zdroj: archiv autora

Proběhlo též porovnávání průběhů rychlostí s jinými zdroji. Při porovnávání s daty z indukčních detektorů bylo zjištěno, že rychlosti při dopravní špičce si odpovídají, a při útlumu dopravy večer či nárůstu v začátku ranní špičky se rychlosti udávané oběma zdroji rozcházejí. Tento jev je způsoben pravděpodobně výchylkou modelu při prudké změně intenzit. Lze si toho všimnout na následujícím průběhu



Obrázek 5- Porovnání rychlostí FCD (modře) s rychlostí dle detektorů RLTC na SOKP(červeně) a jejich rozdíl (zeleně)
zdroj: archiv autora

3.2.2 Časový rámec

Hodnoty jsou poskytovány pro každý časový interval o délce jedné minuty a jedná se o průměr spočítaný modelem na základě odeslaných dat.

Při analýze byl zjištěn časový posun jedné hodiny. Z toho vyplývá nutnost věnovat pozornost času u tohoto datového zdroje a zejména dávat pozor na časová pásma. Dále také byla zjištěna neaktualizovanost dat. V částech dne s nižší intenzitou dopravy jsou data aktualizována po delší době, například deseti minutách, vzhledem k nedostatku zdrojových dat.

Zpětná dostupnost dat není tímto zdrojem omezována.

3.2.3 Prostorový rámec

Data z plovoucích vozidel poskytují údaje o stavu většiny silniční sítě v ČR, která je rozdělena na asi 25 000 jednosměrných segmentů s průměrnou délkou cca 2 km. Tyto segmenty jsou definovány službou RDS-TMC poskytující dopravní informace na silniční síti v rámci Jednotného systému dopravních informací pro ČR (JSDI).

Z teoretické analýzy také vyvstává potřeba prozkoumat kvalitu dat například v tunelech, kde je předpoklad zhoršené kvality z důvodu zhoršených podmínek pro zisk GPS souřadnic.

3.2.4 Zdroje a vlastnosti dat

V rámci výstupů projektu RODOS existují 3 možné způsoby pro získání potřebných datových výstupů DMM. Všechny byly v rámci analýzy vyzkoušeny a prověřeny z hlediska uživatelské i technické funkčnosti. Jedná se o následující rozhraní:

- webové grafické (mapové) rozhraní – RODOS Web View
- webové rozhraní s využitím protokolu SOAP – RODOS Web Services
- samostatná aplikace – RODOS Data

Mapové rozhraní i samostatná aplikace se po analýze jeví jako nepoužitelné, ať už z hlediska špatné stability i chybějících funkcí při práci s daty. Webové rozhraní s protokolem SOAP se jeví jako nejideálnější zdroj dat.

Vzorek dat

Data v jazyce XML jsou strukturována do uzlů „node“, kde jeden uzel představuje údaje za jednu minutu v jednom TMC segmentu. V tomto časovém a prostorovém měřítku poskytuje údaje již popsané ve věcném rámci.


```
<node>

<DateTime>2015-08-01T01:00:00Z</DateTime>
<TmcId>TS01257F10378</TmcId>
<LastUpdate>2015-08-01T00:51:41Z</LastUpdate>
<FreeFlowStatus>
        <Speed>114</Speed>
        <TravelTime>36</TravelTime>
</FreeFlowStatus>
<NormalStatus>
        <Speed>114</Speed>
        <TravelTime>36</TravelTime>
        <Delay>0</Delay>
</NormalStatus>
<CurrentStatus>
        <Speed>97</Speed>
        <TravelTime>42</TravelTime>
        <Delay>6</Delay>
</CurrentStatus>
<Reliability>100</Reliability>
<Load>FreeFlow</Load>

</node>
```


Od vzorku se odvíjí i otázka velikosti těchto dat. Velikost takového vzorku v textových znacích je necelých 0,5kB. Uložíme-li tímto způsobem data nasbíraná za měsíc z každého dostupného TMC segmentu, budeme potřebovat úložiště v řádu stovek GB.

3.3 Další zdroje

Při pátrání po zdrojích, u kterých připadá v úvahu, že by mohly být skladovány z důvodu budoucí využitelnosti, bylo zjištěno jejich nepřeberné množství. Podrobná analýza veškerých zdrojů není v rámci této práce možná z časových důvodů.

Bylo provedeno rozdělení zdrojů podle věcného rámce dat bez ohledu na rámec časový i prostorový.

3.3.1 Ostatní data popisující provoz

Již byly popsány dva známé a dostupné zdroje dat popisující stav silničního provozu. Projekt RODOS, jenž by se díky podobné tematice, kterou se zabývá, dal označit jako konkurenční projekt pro ústavní systém databází. Využívá pro DMM jak plovoucí data, tak detektory TSK. Nadále také využívá dat z mýtných bran na mýtem zpoplatněných komunikacích.

Data z detektorů mýtného systému dle webového rozhraní projektu RODOS uvádí pětiminutovou intenzitu dvou-, tří- a čtyřnápravových vozidel. Tyto data jsou však dle tohoto webového rozhraní souhrnná za všechny jízdní pruhy v daném směru.

Vedle těchto dat jsou na většině mýtných bran vybraných dálnic (D1, D2, D5) umístěny i detektory ASIM. Ty produkují data taktéž pětiminutových intenzit. Jsou rozděleny dle šesti kategorií vozidel:

- Osobní automobily a motocykly
- Nákladní automobily
- Osobní automobily s přívěsem
- Nákladní automobily s přívěsem
- Autobusy
- Neidentifikovatelná vozidla

Na rozdíl od dat z detektorů mýtného systému jsou data z detektorů ASIM rozdělena pro každý jednotlivý jízdní pruh zvlášť. Princip detekce je však jiný – zatímco mýtný systém počítá nápravy vozidel, detektory ASIM využívají kombinace technologií mikrovln, ultrazvuku a pasivního infračerveného záření – toto přináší větší podrobnost, ale zároveň

též obtíže při porovnání těchto dat. Další nevýhodou je absence detektorů ASIM na zbylých mýtných branách. Přístup k datům z detektorů je (dle autorovy zkušenosti ze stáže) možný také pomocí databázových dotazů online. Kontaktní osobou tohoto přístupu je Ing. Kristýna Cikhardtová (FD, Eltodo).

Dalším zdrojem popisující stav provozu na komunikacích může být portál společnosti Google. Ten využívá anonymizovaných dat o pohybu mobilních telefonů od mobilních operátorů či dat z aplikace Waze pro mobilní telefony. Aplikace Waze taktéž kvalitně popisuje infrastrukturu z hlediska neplánovaných uzavírek či jiných omezení.

3.3.2 Geografická data popisující infrastrukturu

Budeme-li se chtít tedy více věnovat infrastruktuře, v nabídce potenciálních zdrojů máme několik možností.

- Český úřad zeměměřický a katastrální provozuje Základní bázi geografických dat (ZABAGED®). Tento produkt je pro zamýšlenou databází důležitý zejména z hlediska prostorového popisu dopravní infrastruktury. Jedná se o linie a uzly silniční i železniční sítě.
- Podobný geografický informační systém je Jednotná dopravní vektorová mapa. Ta je pod hlavičkou Ministerstva dopravy a obsahuje například statistiku nehodovosti na silniční síti či intenzity dopravy jako výstup sčítání dopravy.
- Jednotný systém dopravních informací pro ČR (JSDI) je systém vytvořený za účelem sběru a redistribuce dopravních informací shromažďovaných Národním dopravním informačním centrem (NDIC) za účelem poskytování aktuálních dopravních dat z celého území ČR sjednocených na jediném místě. Systém je komplexní a zahrnuje: dopravní uzavírky, stavební omezení, aktuální omezení v důsledku nehod, zobrazení proměnných tabulí i meteorologická upozornění.
- Známý je též geografický informační systém v komerčním sektoru – Central European Data Agency (CEDA)
- Volně dostupné mapy OpenStreetMap poskytují taktéž popis infrastruktury. Jsou založeny na nekomerčním dobrovolném přístupu k tvorbě map. Obsahují data celosvětově.

3.3.3 Statistická data s vazbou na dopravu

Statistická demografická data mohou významně souviset s mobilitou. Dopravci jsou povinni sdělovat státním orgánům přepravní výkony a další data. Všechna tyto data však nemusí být snadno a veřejně dohledatelná. Pokud jsou zveřejněná, většinou je tomu tak v podobě tabulek umístěných v dokumentech a nikoliv databázově. Mezi největší integrátory statistických demografických dat patří bezesporu Český statistický úřad (ČSÚ). Další instituce jsou například Ministerstvo dopravy či Ministerstvo vnitra.

Ročenky související s dopravou však produkuje mnoho institucí. Může se jednat o:

- České dráhy
- Drážní inspekce
- dopravní podniky měst
- Drážní úřad
- Policie ČR
- Ředitelství silnic a dálnic
- Státní fond dopravní infrastruktury
- technické správy komunikací ve městech

Například ročenka Ministerstva dopravy obsahuje:

- ekonomické ukazatele vývoje národního hospodářství a dopravního sektoru
- údaje o dopravní infrastruktuře
- údaje o vozovém parku
- údaje o přepravě osob i zboží
- údaje o nehodách v dopravě
- vliv dopravy na životní prostředí

3.3.4 Data specifických druhů dopravy

Tato data jsou taktéž rozmanitého charakteru, ale nejsou integrována do celků. Může se jednat o databáze, které využívají soukromé firmy či veřejné instituce zabývající se nějakým specifickým oborem dopravy. Tyto data tedy vyloženě nemají statistický charakter. Například u dopravních podniků lze hovořit o intenzitách přepravených osob s podrobným časovým měřítkem (na linku autobusu, za hodinu v metru atd.) či o datech

pozic dopravních prostředků, ze kterých lze vyčíst mnoho podstatných informací. Podobně zde mohou být data o zpoždění vlaků.

Kromě dopravních podniků se může jednat o data o aktivitě taxislužeb či průjezdu IZS dopravní infrastrukturou města. O nich však zatím není nic známo

3.4 Přístup k zjišťování dat

Dat je nepřeberné množství a jsou dva možné způsoby jejich analýzy.

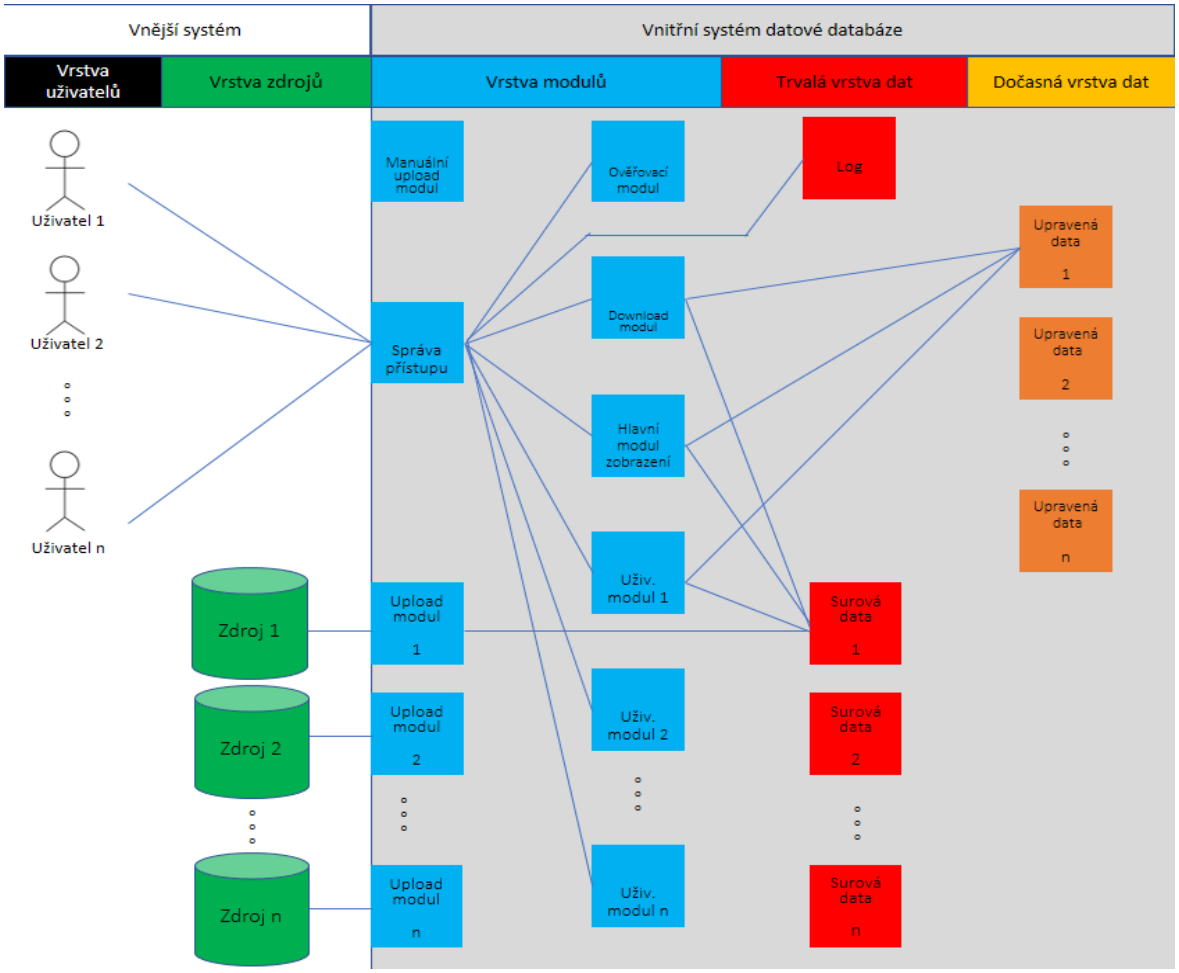
Prvním způsobem je zjistit detaily o co největším množství dostupných dat a z nich vyvíjet výstupy. Nevýhodou tohoto způsobu je vyžadované značné úsilí bez zaručení okamžitého dosáhnutí očekávaných výsledků. Po dlouhotrvajících analýzách vzniká jeden komplexní výstup, jeho kvalitu ale nelze odhadnout.

Druhý způsob je definování výstupu a pátrání pouze po užitečných zdrojích. Jedná se tedy o analýzu možných vstupů s ohledem na jasně definovaný výstup. Tato metoda je charakteristická pro dlouhý postupný vývoj.

4 Systémový popis databáze dat a metadat

Po analýze zdrojů je potřeba popsat navrhovaný systém datové databáze. Ten zde bude pro uvedení představy rozebrán. Systém byl rozčleněn na 4 typy objektů a interakce mezi nimi. Prvním typem jsou uživatelé – ti se dají označit za vnější prvek. Druhému typu vnějších prvků – zdrojům, se věnovala předchozí kapitola. Dvěma vnitřními typy prvků jsou data a aplikace neboli moduly.

Z následujícího obrázku plyne, že všechny tyto 4 typy objektů mají více zástupců rozdělených do několika druhů. Obrázek také ukazuje, že se nejedná o systém s konečným počtem objektů. Interakce mezi objekty jsou pouze naznačené z důvodu komplexnosti. Podrobnější funkce jsou graficky znázorněné na dalších obrázcích.



Obrázek 6 - Náhled celkového systému datové databáze zdroj: archiv autora

4.1 Uživatelé

Uživatel je člověk, který systém jakkoliv používá. Je navrženo několik druhů uživatelů, jejichž popis, pravomoci a zodpovědnosti jsou v této podkapitole shrnuty. Toto rozdělení se vytváří pro odlišení oprávnění systém používat.

- Admin

  Administrátor zodpovídá za technický stav a funkčnost systému, tedy jej spravuje. Má proto všechna oprávnění. Jako jediný může do systému zasahovat – vytvářet tabulky dat, upravovat a mazat. V jeho kompetenci je též spravovat aplikační vrstvu, tedy vkládat a mazat moduly popsané v kapitole 4.3.

- Editor

  Editor na rozdíl od administrátora zodpovídá za obsah databáze. Jako jediný schvaluje data z dočasné vrstvy databáze do vrstvy fixní, čímž je zpřístupňuje pro ostatní uživatele. Také jako jediný může do trvalé vrstvy rovnou vkládat data.

- Výzkumník

  Výzkumník je osoba, která má oprávnění data zobrazit a stáhnout. Pokud data produkuje (ať už manuálně či modulem), může tak činit pouze do dočasné vrstvy databáze.

- Student

  Studentem rozumíme kteréhokoliv studenta. Ten má přístup pouze k metadatům dat, či zpřístupněným datům sloužícím k výuce. Pokud chce zobrazit a stahovat jiná data, musí uvědomit editora či administrátora a požádat jej o zpřístupnění.

- Nepřihlášen

  Kdokoliv, kdo k systému bude přistupovat bez přihlášení, se stává nepřihlášeným uživatelem. Může to být kdokoliv z široké veřejnosti. Tato role umožňuje pouze zobrazit vybraný veřejný obsah, například zjednodušené grafické výstupy.

- Konkrétní zákazník

  Případným zákazníkem bude někdo, kdo si žádá vidět svá zobrazení tak, aby jej neviděla široká veřejnost. Typickým představitelem může být řešitelský tým konkrétního projektu, či město žádající o dopravní datovou expertízu.

4.2 Data a metadata

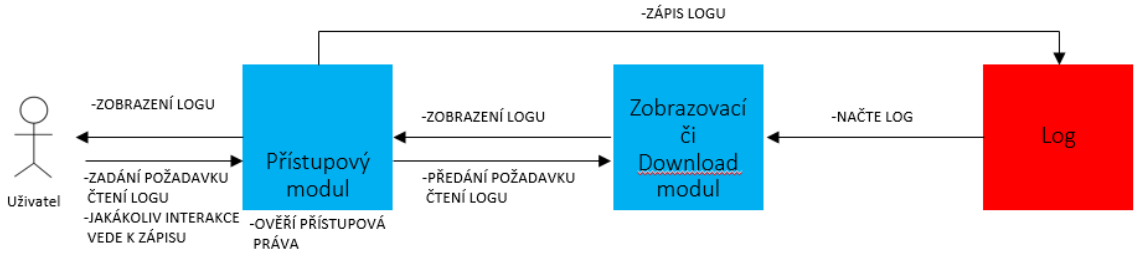
Každá uložená data mají i svá uložená metadata, tedy data a metadata vždy tvoří pár. Je třeba věnovat pozornost přístupovým právům studenta. Ten má možnost zobrazit metadata ke všem datům v trvalé části databáze.

Databáze je rozdělena na trvalou a dočasnou část. Toto rozdělení vzniklo proto, aby dočasná, zatím neschválená data produkovaná aplikačními moduly nebyla nedopatřením považována za věrohodná a také aby přístupy k těmto datům nezatěžovaly ověřená data v trvalé části. V trvalé části jsou uložena ověřená data, zatímco dočasná část je jakási pracovní složka.

Surová data z ověřených zdrojů jsou uložena v trvalé části. Jsou zde také již zpracovaná data, ale až po odsouhlasení editora. Do doby, než editor odsouhlasí nahraná data, jsou uloženy v dočasné části databáze a přístup k nim má pouze editor, administrátor a jejich vlastník. Ostatní uživatelé je vidět nemohou, pokud jim administrátor se souhlasem vlastníka neudělí přístupová práva.

Log je zvláštní typ dat se záznamy veškerých přístupů uživatelů. Přesněji je zde uvedeno kdo, kdy, odkud do systému vstoupil a co v něm dělal. Log je důležitý pro možnost kontroly žádanosti dat. Při zjištění většího zájmu o nějaká data se může dát přednost vývoji těmto datům. Log také upotřebí vyučující předmětu UITS při kontrole, zda

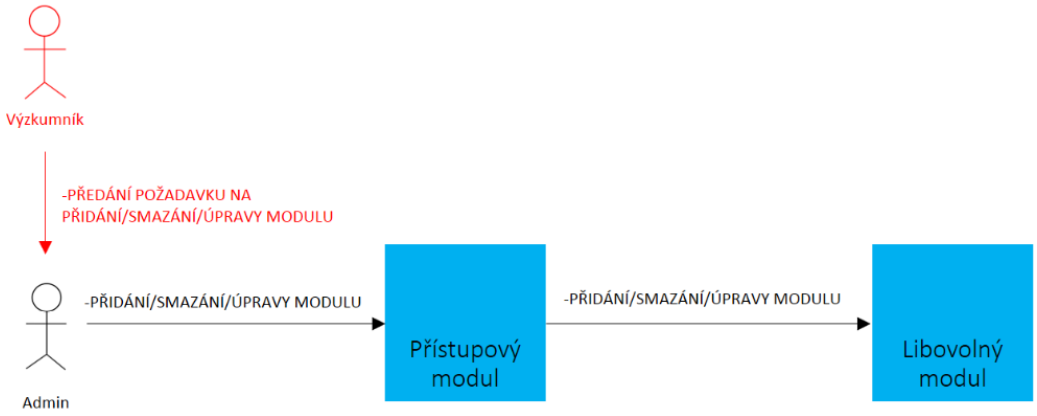
každý student pracoval na zpracování svých dat. Proces zápisu a čtení vystihuje následující obrázek:



Obrázek 7 - Funkce logování
zdroj: archiv autora

4.3 Moduly

Moduly neboli aplikace jsou spolu s databází vnitřními prvky databázového systému. Jde o jakékoliv aplikace či mechanismy spolupracující s databází. Rozděleny jsou do pěti druhů. Jejich správu má na starosti administrátor dle následujícího obrázku.



Obrázek 8 - Správa aplikací
zdroj: archiv autora

4.3.1 Přístupový modul

Přístupový modul je jediný systémový modul. Řídí vstupy uživatelů do systému na základě jejich oprávnění. Uživatelé z ČVUT se přihlašují do systému ČVUT heslem Oprávnění si představíme jako matici, kde na svislé ose jsou vyznačeni všichni uživatelé a na ose

vodorovné všechny činnosti, které lze v systému se všemi prvky dělat. Hodnoty v této matici pak rozhodují o přístupu uživatele.

role	uživatel	objekt / akce	log		dočasná data výzkumníka 1				...		
			zobrazit	smazat	zobrazit	upravit	smazat	ověřit	...	...	...
admin	admin 1		A	A	A	A	A	A			
editor	editor 1		A	N	A	A	N	A			
	editor 2		A	N	A	A	N	A			
výzkumník	výzkumník 1		N	N	A	A	N	N			
	výzkumník 2		N	N	N	N	N	N			
	výzkumník 3		N	N	N	N	N	N			
	výzkumník 4		N	N	N	N	N	N			
...	...										
	...										

Tabulka 2 - Náznak matice přístupů
zdroj: archiv autora

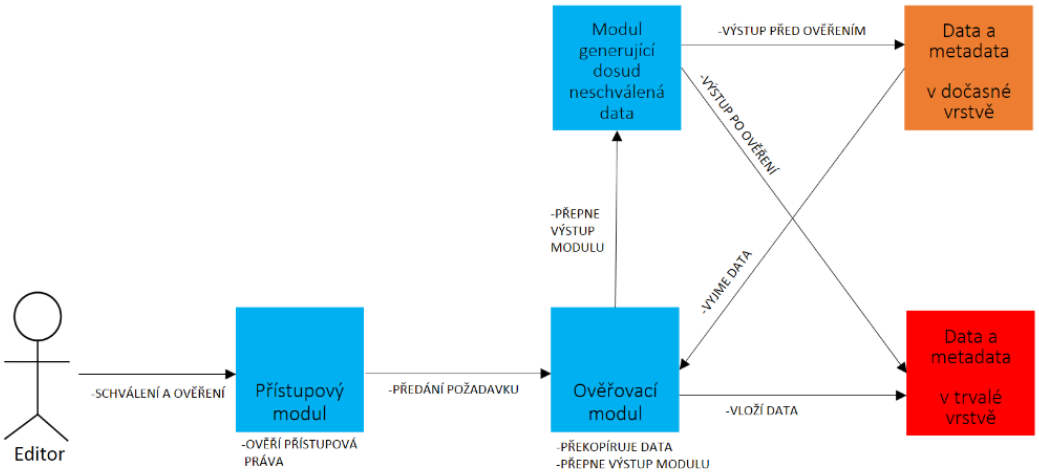
S každým novým prvkem databáze je potřeba rozšířit tuto matici přidáním činností týkajících se nového prvku. Pokud půjde o data, bude se jednat o činnosti zobrazování, přidávání, úpravy a mazání. S každými novými daty se tedy musí definovat kdo k nim, jak může přistupovat. Obdobně je to s novými aplikačními moduly.

Rovněž s každým novým uživatelem je třeba aktualizovat matici přidáním uživatele a nastavením jeho přístupových práv. Tomu pomůže přidělení základní role např. studenta a následné upravení přístupu ke konkrétním datům např. datům pro výuku či projekt.

Druhou funkcí přístupového modulu je zápis logu čili přehledu o uživatelských vstupech a činnostech.

4.3.2 Ověřovací modul

Ověření dat si lze představit jako aplikaci, která zkopíruje z dočasné vrstvy databáze ověřovaná data a vloží je do části trvalé. Následující obrázek popisuje tento proces.



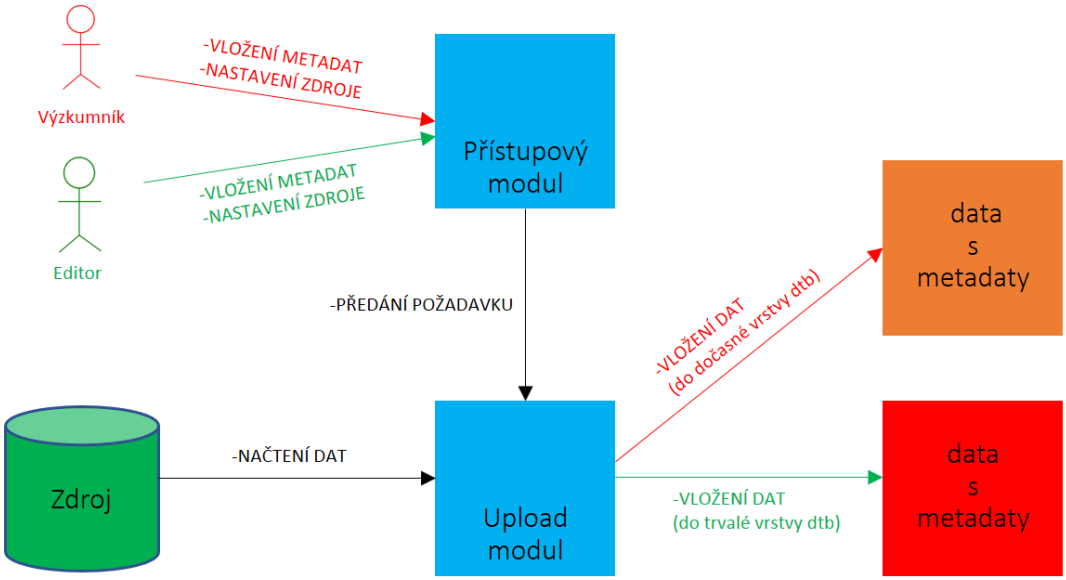
Obrázek 9 - Proces ověřování
zdroj: archiv autora

Editor po prohlédnutí a expertíze ověřovaných dat zadá požadavek na jejich ověření. Pouze on je k tomu určený a oprávněný zároveň. Aplikace po jejím spuštění zkopíruje data z dočasné vrstvy databáze do trvalé a rovněž nastaví výstup aplikace, která data ukládá tak, aby je nyní ukládala do trvalé části.

4.3.3 Upload moduly

Pro nahrávání dat do databáze z externích zdrojů slouží upload moduly. Jejich konkrétní podoba se odvíjí od druhu zdroje. V důsledku rozmanitosti zdrojů a dat bude tedy mít skoro každý zdroj svůj vlastní upload modul.

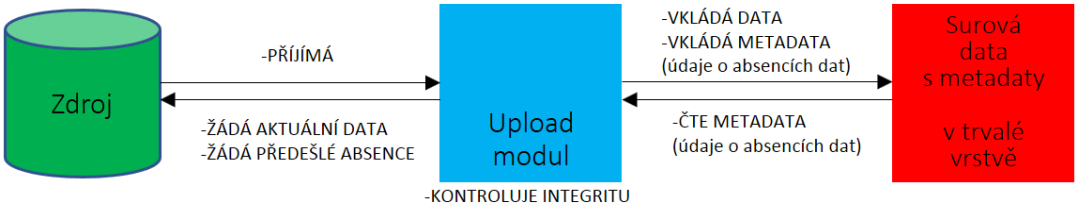
Bude-li potřeba nahrát data manuálně, tedy zásahem uživatele, proběhne manuální proces nahrávání dat, který je zobrazen na následujícím obrázku.



Obrázek 10 - Manuální upload
zdroj: archiv autora

Administrátor nejprve vkládá tabulku, čímž vyhradí prostor pro data. Editor či Výzkumník pak následně pomocí tohoto modulu načte data ze zdroje a vloží je do tabulky (trvalé či dočasné dle role).

Pokud se jedná o automatický proces, i v takovém případě musí admin na začátku vyhradit prostor. Poté následují další události dle následujícího schématu:



Obrázek 11 - Automatický upload
zdroj: archiv autora

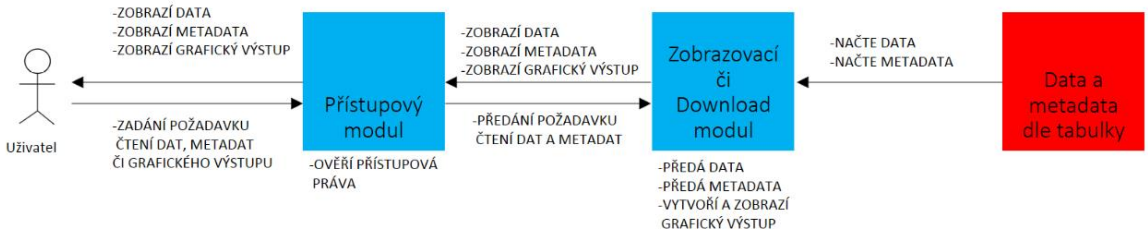
Nejprve upload modul žádá zdroj o data a ten mu je poskytne zpět. Modul si zkontroluje, zda nedošlo k chybě spojení a přijmul skutečně všechna data, která si zažádal. Následně je vloží do databáze spolu s metadaty o kompletnosti. V metadatech zdroje lze dohledat, zda jsou data kompletní. To si kontroluje i upload modul. V případě, že nějaká data chybí,

připojí k další žádosti o data žádost o chybějící data. Upload moduly tedy provádějí prvotní kontrolu dat pro případné dodatečné nahrání. V metadatech jsou tyto nedostatky uloženy. Kontrola datové integrity funguje na základě počtu známých detektorů a časových intervalů, ve kterých poskytují data.

4.3.4 Zobrazovací a Download moduly

Opakem upload modulů jsou download moduly a zobrazovací moduly. Oba slouží k distribuci dat směrem k uživateli. Uživatel tak žádá download modul o data či zobrazovací modul o zjednodušený grafický výstup.

Zjednodušené vybrané výstupy slouží hlavně studentům, veřejnosti či zákazníkům. Ostatní uživatelé mohou skrze zobrazovací modul s daty pracovat, tj. stahovat či zobrazovat je



Obrázek 12 - Zobrazovací a download modul
zdroj: archiv autora

Z předchozího obrázku plyne, že pro zobrazení i download je typický směr pohybu dat k uživateli.

Následující tabulka vyjadřuje doporučená oprávnění zobrazování.

objekt	činnost	upřesnění	role					
			admin	editor	výzkumník	student	anonym	zákazník
data metadata	zobrazení a stažení	dočasná	Ano	Ano	Jen vlastní	Ne	Ne	Ne
		fixní	Ano	Ano	Ano	Všechny metadata Data jen výuková	Ne	Ne
	přidávání	dočasná	Ano	Ano	Jen vlastní	Ne	Ne	Ne
		fixní	Ano	Ano	Ne	Ne	Ne	Ne
	úpravy a mazání	dočasná	Ano	Ne	Ne	Ne	Ne	Ne
		fixní	Ano	Ne	Ne	Ne	Ne	Ne
zjednodušené výstupy	zobrazení	veřejné	Ano	Ano	Ano	Ano	Ano	Ano
		na zakázku	Ano	Ano	Ne	Ne	Ne	Ano
log	zobrazení		Ano	Ano	Ne	Ne	Ne	Ne

Tabulka 3 - Oprávnění skupin uživatelů k zobrazování dat a metadat
zdroj: archiv autora

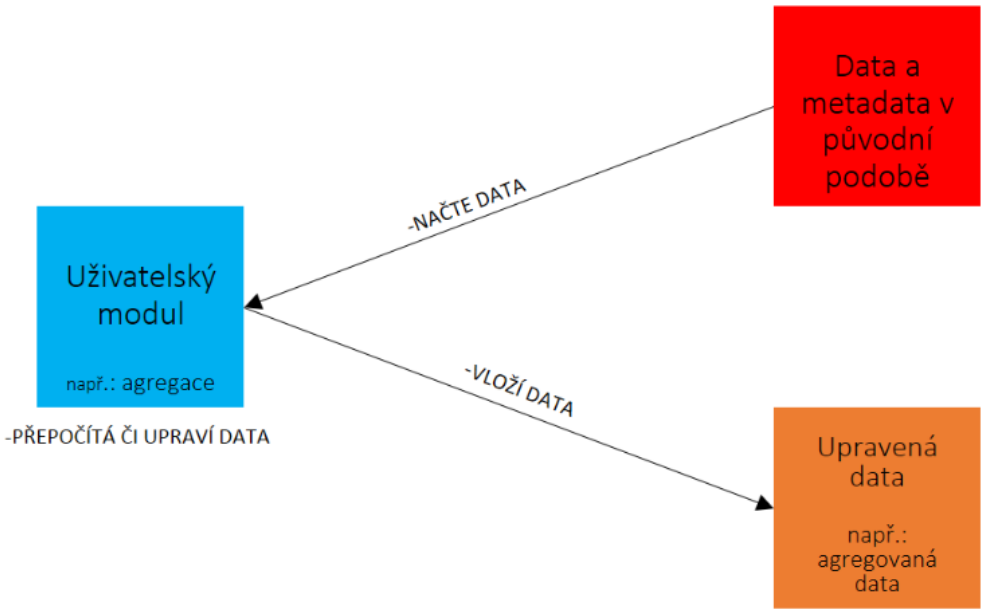
Pro pokročilé stahování dat musí download modul umožňovat všechny možnosti zobrazení, tedy poskytnout soubory typu CSV či XML i vygenerovat SQL dotaz na základě libovolných časových i prostorových preferencí uživatele.

Již na základě uživatelských požadavků byla naznačena problematika datové integrity a nutnost věnovat ji pozornost. Kontrolu datové integrity ve smyslu objemu dat kontrolují příslušné upload moduly. Datová integrita ve smyslu odlehlých hodnot vyžaduje detailní analýzu daných dat. Uživatelovi však může pomoci prvopohledová analýza dat. Tu zajistí zobrazovací modul, který uživatelovi zobrazí statistické hodnoty (počet hodnot celkem či počet hodnot v daném intervalu hodnot, střední hodnotu a rozptyl, …) či zobrazí grafický popis dat v podobě histogramu či krabicového grafu.

Dále zobrazovací modul tvoří mapový metadatový přehled zdrojů dat pro výzkumníky a studenty. Tvoří i vybrané zobrazení pro veřejnost.

4.3.5 Uživatelské moduly

Uživatelských modulů bude nejvíce. Ve zkratce se jedná o moduly, jenž umožní výzkumníkům zpracovávat data. Může se jednat o složité matematické modely, od agregace po predikci. Pokud ještě modul neprošel ověřovacím procesem, múže generovat data jen do dočasné části databáze.



Obrázek 13 - Uživatelské moduly
zdroj: archiv autora

Závěr

Systém databází zahrnuje databázi dat, dokumentů i znalostí. Využití tohoto konceptu přinese mnoho výhod při výuce i výzkumu.

Sběr uživatelských požadavků a rovněž i vizí a myšlenek probíhal formou svolaného brainstormingu se skupinou budoucích uživatelů a poté pomocí dalších individuálních schůzek pro upřesnění detailů.

Byly podrobněji analyzovány dva zdroje dopravních dat. Jedná se o data z plovoucích vozidel dostupná v rámci projektu RODOS a také data z pražských detektorů TSK. Analýza se specializovala na strukturu a prvotní odhalené vlastnosti dat a jejich zdrojů. Byla provedena prvotní rešerše ostatních možných zdrojů v podobě jejich rozdělení a několikačetném jmenování. Byly také naznačeny postupy při rozvoji systému ve smyslu postupnému přibývání datových zdrojů.

Byl proveden základní systémový návrh pomocí dekompozice na objekty a jejich interakce a procesy. Návrh vycházel z uživatelských požadavků.

Tato práce shrnula motivace, vize, cíle a základní uživatelské požadavky. Pomocí osobních schůzek, analýzy datových zdrojů a čerpání informací z odborných příruček byl vytvořen podklad shrnující myšlenky plánovaného budoucího nástroje.

V dalším vývojovém kroku je potřeba sestavit konkrétní technické řešení datové části a nadále provádět analýzy dalších zdrojů dat. Bylo by též vhodné se současně věnovat detailněji konceptům wiki a skladování dokumentů.

Autor věří že jím vytvořená publikace sdělí myšlenky a vize všem, kteří se budou tématem dopravních databází nadále zabývat a bude inspirativní.

Seznam literatury

[1] ROESS, Roger P., Elena S. PRASSAS a William R. MCSHANE. *Traffic engineering*. 4th ed. Upper Saddle River, NJ: Pearson, c2011. ISBN 0-13-613573-0. kapitola 8

[2] ČARSKÝ, Jiří. *Bílá kniha ČVUT FD*. Praha, 2017.

[3] KREJČÍ, Jakub. *Detektory TSK*. Praha, 2016.

[4] *Strategické detektory TSK* [online]. Eltodo [cit. 2017-06-25]. Dostupné z: stratdet.eltodo.cz

[5] KABELKA. *TRANSMITTER 3 komunikační protokol*. Praha, 2016.

[6] HRUBEŠ, Pavel, Martin LANGR, Jiří RŮŽIČKA a Jakub KREJČÍ. *Porovnání nových optimalizačních metod s reálnými daty na vybrané oblasti s daty z DMM*. Praha, 2017.

[7] KOCYAN, Tomáš. *Rozhraní pro poskytování dat systému RODOS*. 2016.

Seznam obrázků

Obrázek 1- Schéma systému databází; zdroj: archiv autora 9

Obrázek 2 - Myšlenková mapa jako náznak rozdělení zdrojů; zdroj: archiv auora 20

Obrázek 3 - Mapa strategických detektorů Eltodo zdroj: stratdet.eltodo.cz 24

Obrázek 4- Histogram rychlostí FCD zdroj: archiv autora..... 28

Obrázek 5- Porovnání rychlostí FCD s rychlostí dle detektorů RLTC na SOKP 28

Obrázek 6 - Náhled celkového systému datové databáze zdroj: archiv autora..... 35

Obrázek 7 - Funkce logování zdroj: archiv autora 38

Obrázek 9 - Proces ověřování zdroj: archiv autora..... 40

Obrázek 10 - Manuální upload zdroj: archiv autora 41

Obrázek 11 - Automatický upload zdroj: archiv autora 41

Obrázek 12 - Zobrazovací a download modul zdroj: archiv autora 42

Obrázek 13 - Uživatelské moduly zdroj: archiv autora 44

Seznam tabulek

Tabulka 1- Přehled detektorů TSK; zdroj: archiv autora..... 21

Tabulka 2 - Náznak matice přístupů zdroj: archiv autora..... 39

Tabulka 3 - Oprávnění skupin uživatelů k zobrazování dat a metadat 43