

České vysoké učení technické v Praze
Fakulta stavební



**155UZPR Úvod do zpracování prostorových
dat**

**Nasazení vektorových dlaždic při
tvorbě katastrální mapy**

Bc. Linda Kladivová, Bc. Jana Špererová, Bc. Lukáš
Kettner, Bc. Martin Hulín
3. februára 2020

Obsah

1	Úvod	2
2	Popis a rozbor problému	3
2.1	Vektorový a rastrový formát	3
2.2	Súčasný stav mapového serveru Marushka	4
2.3	Možnosti generování vektorových dlaždic	5
3	Popis a formát vstupních dat	7
4	Popis a formát výstupních dat	8
4.1	MBTiles	8
4.2	Formát PFB ("Protocolbuffer Binary Format")	8
4.3	Geojson	8
5	Pracovní postup	10
5.1	Stážení shapefilů katastrální mapy	10
5.2	Import shapefilů do PostGIS databáze	10
5.3	Export dat z formátu shapefile do formátu geojson	11
5.4	Tvorba vektorových dlaždic	11
5.5	Vybalení vektorových dlaždic	12
5.6	Vykreslení vektorových dlaždic	12
5.6.1	Leaflet	12
5.6.2	Leaflet Vector Grid	12
6	Záver, možné a neriešené problémy	13

1 Úvod

Cílem této semestrální práce bylo, ze shapefilových vektorových dat katastrálních území vytvořit vektorové dlaždice a najít nejvhodnější nastavení pro jejich prezentaci na webových prohlížečích. Vektorové dlaždice, jsou ve své podstatě rozřezané vektorové mapy na jednotlivé úseky podle nějaké mřížky. A každá dlaždice je pak samostatným článkem. Například tedy, že už na webu nemusím načítat celou velkou mapu, ale načtu pouze jednu konkrétní dlaždici, která odpovídá lokalitě, kterou potřebuji. Tím, že vektorové dlaždice jsou pouze část celku, mají mnohem menší objem dat a tedy jsou méně náročné na zpracování. Vytváření vektorových dlaždic nemá zatím v dnešní době žádný standart pro jejich vytváření. Jedinou výjimkou je Mapbox, který specifikuje nějaké předpoklady pro jejich vytváření. Pro vytvoření vektorových dlaždic v této práci, jsme se rozhodli pro použití nástroje Tippecanoe. Nástroj Tippecanoe jsme vybrali z důvodu, že je to opensource nástroj, který je veřejně dostupný např. na GitHubu. Našimi vstupními daty pro tuto funkci byla data poskytnutá ČUZK ve formátu shapefile. Tato data byla převedena na geojson, ze kterého byly vygenerovány jednotlivé vektorové dlaždice ve formátu pbf a ty pak nahrány na webový server.

2 Popis a rozbor problému

2.1 Vektorový a rastrový formát

Každý z formátov publikácie priestorových dát má svoje výhody a nevýhody. Je potreba dobre zvážiť umýsel, možnosti a výsledný cieľ publikácie. Rastrové dlaždice majú v porovnaní s vektorovými jednoduchšiu dátovú štruktúru a spracovanie rastrových dát je vo väčšine prípadov jednoduchšie. Veľkou nevýhodou dát v rastrovom formáte sú obmedzené možnosti práce s dátami, kedy pri porovnaní s vektorovými dátami pracujeme s obrazom javu a nie konkrétnymi geoprvkami. Ďalšou nevýhodou rastrových dlaždíc je ich veľkosť. Vzhľadom na využitie práce, riešenie má byť aplikované pre dáta celého územia Českej republiky, je práve veľkosť kľúčovým faktorom, ktorý hovorí v prospech vektorových dlaždíc. Z hľadiska úložiska vektorové dlaždice zaberajú kapacitne niekoľkonásobne menší (väčšinou) priestor úložiska. Z tohto faktu plynie ďalšia výhoda pri použití vektorových dlaždíc, úspora dátového toku na klientovi, vykresľujú sa dlaždice, ktoré sú v aktuálnom pohľade mapy.

Ďalším podstatným plusom oproti rastrovým dlaždiciam je možnosť vektorového formátu niesť topologickú informáciu, ktorú môže užívateľ využiť pri následnej práci s dátami a vykonávaní rozličných analýz. Jedným z príkladov je možnosť vykonávať úpravy priamo na strane klienta, bez nutnosti interakcie so serverom. Táto možnosť je užitočná pri projektoch v oblasti open source mapovania a práci s využitím mapovej služby priamo v teréne.

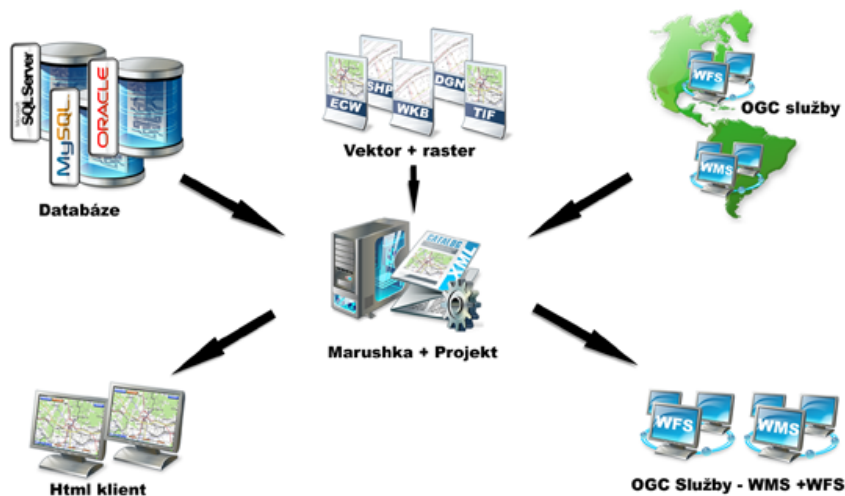
K výhodám vektorových dlaždíc patrí aj kartografická generalizácia. V závislosti na úrovni priblíženia (merítke) je možné nastaviť optimalizáciu zobrazenia vektorových dlaždíc. Jednotlivé geoprvky je možné zlúčiť, zjednodušiť či agregovať, tak aby podrobnosť geometrie a atribútového obsahu dlaždíc odpoveda úrovni priblíženia.

Jednou z ďalších výhod vektorových dátach je možná individuálna štylizácia dát na strane klienta, presne podľa jeho požiadavkov, bez toho aby dáta na serveri museli byť zmené.

V prípade modelového príkladu z praxe. Pokiaľ by sme chceli na mobilnom zariadení, v lokalite so slabším internetovým pripojením vykonať jednoduchú analýzu nad mapou vytvorenou z rastrových dlaždíc, sme takmer bez šance na úspech. V prípade použitia vektorových dlaždíc bude možné túto analýzu vykonať.

2.2 Súčasný stav mapového serveru Marushka

V súčasnom stave mapový server Marushka slúži k uloženiu, poskytovaníu a aktualizácii mapových dát na území Českej republiky. Všetky dáta publikované touto aplikáciou sú uložené v databázy Informačného systému katastru nemovitostí (ISKN) a v databázy Informačného systému územnej identifikácie (ISÚI). V rámci týchto dvoch databází prebiehajú všetky aktualizácie zobrazených prvkov. Na tieto databázy je následne napojená Publikačná databáza a v nej prebiehajú úpravy geometrií v pravidelných dvojhodinových cykloch. Úpravy sa týkajú zväčša transformácie geometrie, aby bolo možné jej optimálnejšie zobrazovanie a publikácia. Celá logika postupu je na základe databázovej úrovne a je realizovaná pomocou PL/SQL skriptov. Takto upravené dáta sú automaticky exportované do WKB súborov. Z týchto preddefinovaných súborov prebieha samotná publikácia dát vo forme rastrových dlaždíc. K aktualizácii katastrálnej mapy dôjde v celom katastrálnom území, pokiaľ sa v ňom objaví zmena.



Obr. 1: Mapový server Marushka

Pre lepšie porozumenie je potrebné objasniť, čo je to WKB súbor. Mapový server Marushka natívne pracuje s týmto vektorovým formátom, Well known Binary (WKB, štandard konsorcia OGC). Tento formát je používaný aj pre implementáciu databázového priestorového úložiska, ktoré je kompatibilné s produktami fy GEOVAP(Geostore), prípadne pre export akéhokoľvek skladu do tzv. file cash.

Nad súbormi WKB je možné vybudovať viacúrovňový priestorový index reprezentovaný štruktúrou R-Tree. Tento index výrazne zefektívni priestorový prístup k jednotlivým geometrickým elementom. Podrobnejšie informácie je možné nájsť na oficiálnej stránke [serveru Marushka](#).

Využitie vektorových dlaždíc na mapovom serveri Marushka by viedlo k zjednodušeniu aktualizácie katastrálnej mapy. Aktualizovala by sa len základná dlaždica, v ktorej sa prvok nachádza. To by vylúčilo nutnosť aktualizovať celé katastrálne územie, tým by sa výrazne znížila časová náročnosť pre aktualizáciu dát a aktuálne dáta na mapovom serveri Marushka by mali používatelia k dispozícii takmer ihneď po úprave.

2.3 Možnosti generování vektorových dlaždíc

V dnešnej dobe je možnosť tvorby vektorových dlaždíc pomocou niekoľkých nástrojov.

Spoločnosť Mapbox vytvorila otvorenú špecifikáciu apbox Vector Tile - MVT. Jedná sa o nástroj, ktorý umožňuje tvorbu mapových štýlov pre vektorové dlaždice nad vlastnými dátami. Základná varianta poskytovania mapových dlaždíc je dostupná zdarma. Spoločnosť Mapbox ponúka aj ďalšie nástroje, ktoré môže užívateľ využiť pri tvorbe mapovej aplikácie využívajúcej ako podklad vektorové mapové dlaždice. Formát MVT je vyvinutý tak, aby požadované údaje boli súvislo v pamäti a aby vrstvy mohli byť bez úpravy údajov pripojené k dlaždici. Každá dlaždica by mala obsahovať aspoň jednu vrstvu, vrstva aspoň jeden prvok a prvok musí obsahovať pole geometrie. Prvky dlaždíc obsahujú ID, atribúty a geometriu. Geometria môže byť definovaná ako bod, línia či polygón. Geometria sa ukladá vo formáte MVT ako pole čísel x a y . Začiatok jednej dlaždice je v ľavom hornom rohu a koniec na dolnom okraji alebo v dolnom pravom rohu (Mapbox 2017).

Ďalšou možnosťou je využitie projektu OpenMap Tiles. Jedná sa o sadu nástrojov pre tvorbu a štylizáciu vektorovej mapy. Tento projekt je nástupcom OSM2VectorTiles projektu. Zahŕňa nástroje pre tvorbu vektorových dlaždíc z OSM (MVT), nástroj pre správu schémy vektorových dlaždíc (popis jednotlivých tématických datových vrstiev a atribútov), nástroj pre predpripravené štýly pre dlaždicové dáta a ich jednoduché prispôbenie. Tento nástroj je možné využiť na počítači s Windows, MAC či Linux operačným systémom, čož je jeho obrovská výhoda napríklad proti nástroju Tippecanoe, ktorý je dostupný len pre Linux.

Často používaným spôsobom je už spomenutý nástroj Tippecanoe. Nástroj

prevádza jednotlivé dáta z formátu GeoJSON, Geobuf alebo CSV do vektorových mapových dlaždíc. Výstupom je formát so špecifikáciou MBTiles. Je ho možné považovať za SQL databázu, ktorá umožní uloženie vektorových mapových dlaždíc v jednom súbore. Tento nástroj bol využitý pri zobrazení vektorových dlaždíc v rámci nášho projektu. Veľkou výhodou bola dostupnosť podrobnej dokumentácie a vzorových príkladov na [GitHub](#).

Pri požiadavku generovania vektorových dlaždíc priamo z databáze Oracle je možné využiť niekoľko postupov. Prvým z nich je generovanie vektorových dlaždíc pomocou Geoserveru. Ten avšak nemá vstavanú podporu pre Oracle databázu. Toto rozšírenie je potrebné doinštalovať. Pri inštalácii je potrebné zohľadniť fakt, že nie všetky požadované časti rozšírenia sú zakomponované v licenčných podmienkach. Preto je potrebné niektoré súbory rozšírenia stiahnuť dodatočne. Podrobný postup inštalácie rozšírenia je možné nájsť na stránkach [GeoServeru](#). Tento nástroj je rozdelený na komerčnú a open source verziu. Komerčná verzia zahŕňa balík nástrojov, ktorý umožní generovať vektorové dlaždice vo formáte GeoJSON, TopoJSON alebo Mapbox. Open source verzia umožňuje generovanie vektorových dlaždíc vo formáte MVT.

Ďalšou možnosťou generovania vektorových dát priamo z Oracle databáze je využitie MapServeru. Postup je rovnaký ako v predchádzajúcom prípade. Je potrebná inštalácia Oracle Spatial. Podrobný postup inštalácie je k dispozícii na webovej stránke [MapServeru](#).

Treťou možnosťou je načítanie dát z Oracle databáze pomocou PostgreSQL a využitie GDAL/ORG data wrapper (datových prevodníkov) a následné dokončenie vykreslenia vektorových dlaždíc v prostredí PostgreSQL.

3 Popis a formát vstupních dat

Našimi počátečními vstupními daty jsou data katastrální mapy ve formátu shapefile rozčleněné po katastrálních územích. Každé k. ú. obsahuje celkem 84 datových položek (tedy 21 shapefilů, jelikož shp je složen ze čtyř souborů). Jedná se o skupiny: bodová pole, budovy, další prvky mapy, hranice parcel, katastrální území, parcely KN, orientační prvky mapy a věcná břemena. Shapefily katastrální mapy jsou přibližným obrazem publikační databáze na ČÚZK. Vzhledem k tomu, že v publikační databázi může být v jedné tabulce i vícero sloupců s různými typy geometrií, neodpovídají názvy shapefilů názvům tabulek v publikační databázi. Aby bylo zřejmé, o jaký konkrétní prostorový typ dat se jedná, jsou jednotlivé shapefily pojmenovány podle svých typů. Jednotlivé typy jsou uvedeny na příkladu vrstvy PARCELY_KN:

- PARCELY_KN_B.shp → body parcel (multipoint)
- PARCELY_KN_DEF.shp → definiční body parcel (multipoint)
- PARCELY_KN_L.shp → linie parcel (multilinestring)
- PARCELY_KN_P.shp → polygon parcel (multipolygon)
- PARCELY_KN_T.shp → text (multipoint)

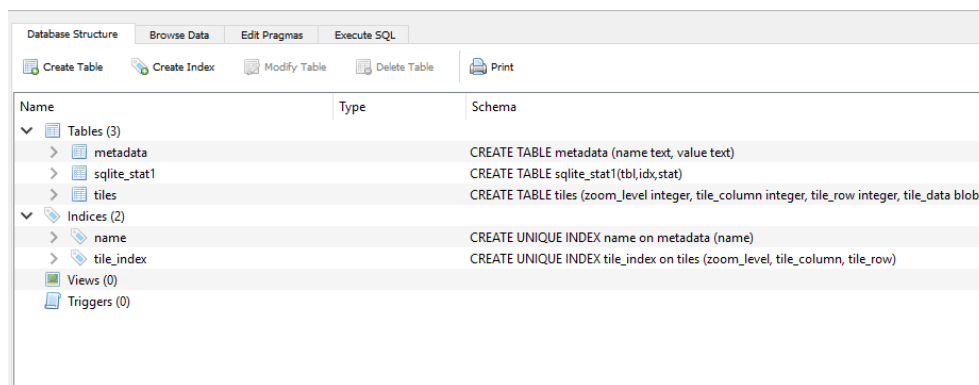
Shapefile je datový formát pro ukládání vektorových prostorových dat, který se skládá z několika povinných a doplňkových souborů. Hlavní soubor .shp obsahuje seznam lomových bodů dané geometrie. Databázový soubor .dbf obsahuje popisné atributy jednotlivých geometrií, které v tomto případě kopírují popisné informace v tabulkách v publikační databázi. Tyto dva jmenované soubory jsou propojeny pomocí dalšího posledního povinného souboru .shx. U všech vrstev je obsažen ještě soubor .prj, který je ve všech případech stejný. Obsahuje definici souřadnicového systému a projekce (EPSG kód 5514).

Mezikrokem této práce je pokus o jednoduché napodobení publikační databáze ČÚZK, jelikož v reálném příkladě počáteční data pro generování vektorových dlaždic nemají charakter shapefilů, ale prostorových sloupců typu SDO_GEOMETRY v Oracle databázi s indexem typu Rtree (stromová datová struktura používaná pro prostorová data, která dělí a seskupuje blízká data do MER - minimal enclosing box). Vzhledem k tomu, že použití nadstavby Oracle Spatial není bezplatně možné, byla zvolena varianta řešení v databázi PostGIS.

4 Popis a formát výstupních dat

4.1 MBTiles

Jde o formát, který slouží pro ukládání velkého množství mapových dlaždic v jednom souboru. V podstatě se jedná o jednoduchou souborovou sqlite databázi s pár tabulkami. Může obsahovat rastrová nebo vektorová data.



Name	Type	Schema
Tables (3)		
metadata		CREATE TABLE metadata (name text, value text)
sqlite_stat1		CREATE TABLE sqlite_stat1(tbl,idx,stat)
tiles		CREATE TABLE tiles (zoom_level integer, tile_column integer, tile_row integer, tile_data blob)
Indices (2)		
name		CREATE UNIQUE INDEX name on metadata (name)
tile_index		CREATE UNIQUE INDEX tile_index on tiles (zoom_level, tile_column, tile_row)
Views (0)		
Triggers (0)		

Obr. 2: Ukázka struktury souborové databáze

4.2 Formát PFB ("Protocolbuffer Binary Format")

Tento formát používají pro ukládání dat OpenStreetMaps.

Např: https://wiki.openstreetmap.org/wiki/PBF_Format
<https://github.com/mapbox/pbf>

4.3 Geojson

Geojson je otevřený formát, který jednoduše prezentuje prostorová data i s jejich atributy. Nejčastěji je využíván pro svoji jednoduchost, snadnou čitelnost a malý objem dat. Díky tomu není tak náročný na zpracování, proto se hodí například pro webové prohlížeče. Tento formát je založen na JavaScript Object Notation (JSON). Předpona geo značí práci s geodaty. Referenční souřadný systém pro použití geojson formátu bývá zpravidla WGS84. Json je formát, který používá textový zápis dat nezávisle na platformě. A lze používat následující typy dat:

string ... textový řetězec (např.: „Hello world“)

number ... číslo (celočíslné nebo reálné) (např.: 0.1e-12 nebo 0.7 nebo -12)

atd.)

boolean ... logická hodnota (true nebo false)

null ... nulová(prázdná) hodnota (zadáváme pouze null)

array ... pole (např.: [1, „vektorové dlaždice“, true] atd.)

object ... objekt (např.: „x“ : 5000, „y“ : 1000 atd.)

V našem případě bylo použit jeden objekt, který se uvnitř větvil na další části a udával např. minimální a maximální možné přiblížení nebo jaké souřadnice jsou výchozí souřadnice pro zobrazení dlaždic.

4.4 Vektorové dlaždice

Výhodou vektorových dlaždic je, že uživatel si zobrazí pouze data z konkrétní lokality, která ho zajímá a nemusí načítat okolní data, která jsou pro něj nepotřebná. A tím pádem ušetří čas při čekání, než by se všechna okolní data načetla. Díky vektorové dlaždici načte pouze konkrétní dlaždici a zbytek dat, jako kdyby v danou chvíli „neexistoval“. Abychom mohli jakoukoli mapu převést na vektorovou dlaždici, nejprve potřebujeme, aby výchozí dataset byl vektorové struktury. Tedy, že každý prvek má svoji geometrickou informaci. Tedy, že každý prvek se skládá z křivek. Pokud máme vektorový dataset připravený, stačí si vytvořit grid, podle kterého daný dataset „rozdělíme“ na jednotlivé vektorové dlaždice.

5 Pracovní postup

5.1 Stažení shapefilů katastrální mapy

Pro automatické stažení dat po katastrálních územích je na stránkách [Services ČÚZK](#) k dispozici zásuvný modul pro hromadné stahování dat KM ve formátu .shp. V Oracle databázi ISKN byl nejprve proveden jednoduchý výběr několika různě velkých skupin k. ú. a jejich export do textových souborů. Po zadání připraveného seznamu byly staženy shapefilly rozčleněné do složek podle kódu k. ú. Nakonec byly staženy shapefilly pro tři různě velké skupiny dat: pro obec Boskovice (52 MB), pro okres Benešov (1 GB, 271 k. ú.) a pro celý liberecký kraj (přes 2 GB, 508 k. ú.). Nejprve však bylo pracováno pouze s nejmenší množinou dat, s pěti katastrálními územími 600822, 608327, 608475, 608483, 785598 spadajícími pod obec Boskovice.

5.2 Import shapefilů do PostGIS databáze

Import shapefilů do PostGISu byl naprogramován automatizovaně v prostředí Windows Powershell (zkráceně PS). Jelikož jsou shapefilly různých datových typů a i jejich import musí proběhnout do různých tabulek v databázi, byl sestaven multidimenzionální řetězec. Všechny složky jsou pak procházeny pomocí for cyklu a data nahrány do databáze přes utilitu **ogr2ogr**.

Při importu bylo řešeno několik problémů. Aby bylo možné ogr2ogr volat relativně, bylo nutné do proměnných prostředí PATH doplnit cestu ke zdrojovému adresáři softwaru QGIS a cestu ke zdrojovému adresáři knihovny PROJ. Důležité je zdůraznit spuštění PS jako administrátor, jinak prostředí systémové proměnné ignoruje.

Další problém, který záhy nastal, byl problém s kódováním shapefilů. V případě těchto dat bylo nutné použít pro import do databáze jiné kódování, než je UTF8 a to LATIN1. Toto kódování bylo nastaveno přímo v PS do proměnné prostředí PGCLIENTENCODING.

Do tabulek bylo přidáváno a nad každou tabulkou byl vytvořen prostorový index. Byl zaznamenán případ, kdy nějaký řádek v datech není datového typu, který má reprezentovat, i přesto že ostatní řádky v shapefilu tento datový typ splňují. Vzhledem k tomu, že to bylo zaznamenáno pouze u pár případů a v této práci nejde ani tak o kompetnost dat, jako spíše o jejich strukturu, takové případy nebyly dále zkoumány. Příkaz ogr2ogr pro import dat byl vytvořen s následující strukturou (proměnné jsou označeny \$):

```
ogr2ogr -append -f PostgreSQL PG:" dbname='pgis_uzpr'
user='uzpr20_a' password='a.uzpr20'
host=' geo102.fsv.cvut.cz '" -a_srs $sour_system $dir
-nlt $typ -lco GEOMETRYNAME=geom
-lco SPATIAL_INDEX=GIST -nln "uzpr20_a.$tabulka"
```

V databázi bylo tedy celkem vytvořeno 21 tabulek (podle názvů shapefilů), do kterých byla naimportována data ze všech katastrálních území. Data je možné si zobrazit v QGISu.

Obrázky dat v QGISu

5.3 Export dat z formátu shapefile do formátu geojson

Pomocí skriptu v Power shellu byly *.shp soubory převedeny na *.geojson soubory, ze kterých se následně tvoří vektorové dlaždice utilitou **tippecanoe**.

Zároveň zde byl změněn souřadnicový systém dat z WGS84 na Web Mercator s kódem EPSG:3857.

Při práci nastalo několik problémů, které bylo potřeba řešit.

- Povolení spouštění skriptů: set-executionpolicy remotesigned
- Nastavení systémové proměnné GDAL_DATA tak, aby ukazovala na adresář se souřadnicovými systémy
- Změna cílového souřadnicového systému na Web Mercator – EPSG:3857

5.4 Tvorba vektorových dlaždic

Cílem této práce je tvorba vektorových dlaždic s různým zadáním parametrů, tak aby bylo zřejmé, jak funguje utilita **tippecanoe**, a jaké nastavení by bylo vhodné použít v případě dat katastrální mapy.

Podporovaný souřadnicový systém vstupních dat je WGS-84 (default) s EPSG kódem 4326 nebo Web Mercator EPSG:3857. Pro druhou variantu je potřeba pomocí vstupního parametru -s upřesnit, že jde o vstupní data ve formátu Web Mercator.

Výsledná souborová databáze ve formátu mbtile je v souřadnicovém systému Web Mercator.

Dlaždice byly vytvořeny příkazem:

```
tippecanoe -o ./scratch/tiles.mbtiles -zg -pk -pC -pS -pt -f  
./UZPR_data/geojson_obec_boskovice/600822_BUDOVY_P.geojson
```

Význam jednotlivých parametrů:

-o Zapsat nové dlaždice do specifikovaného .mbtiles souboru

-zg Automaticky vyber max zoom tak,
aby bylo možné dostatečně rozlíst prvky i jejich detaily

-pk Bez omezení velikosti na 500KB

-pC Nekomprimovat vektorová data ve formátu PBF.

Chyba "Unimplemented type 3" je nejspíše způsobena komprimovanými dlaždicemi

-pS Negeneralizovat linie a polygony při max zoom (pro menší zoom zjednotí)

-f Odstraní výstup.mbtiles pokud již existuje

5.5 Vybalení vektorových dlaždic

Ze souborové databáze ve formátu mbtile se vytvoří struktura malých vektorových dlaždic ve formátu PBF, které si klient načítá. Nemusí tedy stahovat ze serveru do paměti celou databázi mbtile, která může být velmi velká. Posílá dotazy jen na menší soubory pbf, které jsou ve struktuře x, y, z.

5.6 Vykreslení vektorových dlaždic

5.6.1 Leaflet

Umožňuje zobrazit pouze rastrová data.

5.6.2 Leaflet Vector Grid

6 Záver, možné a neriešené problémy

Výsledkom práce je ...

Zoznam obrázkov

1	Mapový server Marushka	4
2	Ukážka štruktúry súborovej databázy	8

Zoznam tabuliek