Úvod do zpracování prostorových dat

Kvalita bydlení ve vybraných částech Prahy

Zimní semestr 2018/2019

Tereza Kulovaná
Markéta Pecenová

Obsah

1 Zadání

1.1 Zadání projektu

Navrhněte a vytvořte tématické vrstvy (např. vodní toky, vodní plochy, lesy, silnice, železnice apod.) na základě dat *OpenStreetMap* a další otevřených zdrojů. Aplikujte testy datové integrity a odstraňte případné nekonzistence v datech. Vytvořte tutoriál - tj. sadu atributových a prostorových dotazů nad databází **pgis_uzpd**.

1.2 Zvolené téma

Jako téma pro semestrální projekt byla zvolena analýza kvality bydlení v Praze. Jelikož je toto téma velmi obecné a obsáhlé, bylo nutné zvolit užší zaměření. Z mnoha podkladů, které byly k dispozici, byly vybrány datové vrstvy, které nějak souvisely s občanskou vybaveností nebo zdravím obyvatel Prahy. Výsledným produktem je sada atributových a prostorových dotazů, které vyhodnocují kvalitu místa pro život na základě zvolených vstupních ukazatelů kvality.

2 Data

V rámci projektu bylo nutné získat potřebná vstupní data, nahrát je do databáze a zajistit jejich konzistenci. Nad těmito daty pak byly následně provedeny prostorové dotazy.

2.1 Data a tématické vrstvy

Data použitá v rámci projektu pochází ze tří zdrojů: ze schématu *ruian_praha*, portálů *Otevřená data* a *Geoportál Praha*.

2.1.1 ruain_praha

Schéma *ruian_praha* je součástí databáze **pgis_uzpd** a je v souřadnicovém systému JTSK.

Vrstvy:

- *adresni_mista* (adresnimista)
- *obvody* (spravniobvody)

2.1.2 Otevřená data

Veškeré datové vrstvy, které byly staženy z portálu *Otevřená data*, mají uvedeného jako poskytovatele *HLAVNÍ MĚSTO PRAHA* a jsou vztažena pouze na území Prahy. Geometrie všech vrstev je reprezentována bodem. Data byla stažena ve formátu geoJSON a až na poslední uvedenou vrstvu byla v souřadnicovém systému WGS84 (S-JTSK nebyl k dispozici). Pouze *Vstupy do metra* byly staženy přímo v S-JTSK.

Vrstvy:

- *detska_hriste* (Dětská hřiště Praha, [Zdroj])
- *zdrav_zarizeni* (Lékárny a zdravotnická zařízení v Praze, [Zdroj])
- *metro* (Vstupy do metra, [Zdroj])

2.1.3 Geoportál Praha

Ze stránek *Geoportál Praha* byla stažena tématická data zachycující rozmístění košů na tříděný odpad a data zaplaveného území Prahy při povodních v roce 2013. Data byla stažena ve formátu geoJSON v systému JTSK. Geometrie rozmístění košů je *bod* a zaplaveného území *polygon*. Pro povodňová data bylo k dispozici ke stažení vrstev mnohem více, například záplavové čáry pro stoletou/padesátiletou/dvacetiletou vodu atd. Pro zjednodušení byla stažena data, který byla ucelená (zachycující jedno období) a nejaktuálnější.

Vrstvy:

- *odpad* (Mapa košů na tříděný odpad, [Zdroj])
- *zaplavy2013* (Záplavové území 2013, [Zdroj])

3 Zpracování dat

3.1 ruian_praha

Adresní místa

Data byla stažena ze schématu *ruian_praha* (vrstva *adresnimista*) do nově vytvořené tabulky *adresni_mista*. Z původních dat byly zkopírovaný pouze sloupečky *kód*, *číslo popisné*, *číslo orientační* a sloupeček s geometrií *geom*.

```
CREATE TABLE adresni_mista AS
    SELECT kod, cislodomovni, cisloorientacni, geom
    FROM ruian_praha.adresnimista
```

Následně byl nad sloupečkem *kód* nastaven primární klíč a nad sloupečkem *geom* prostorový index.

```
ALTER TABLE adresni_mista ADD PRIMARY KEY(kod)

CREATE index adresy_index ON adresni_mista(geom)
```

Na závěr byla provedena kontrola validity geometrie, která vyšla negativní.

```
SELECT kod FROM adresni_mista WHERE NOT st_isvalid(geom)
```

Při zběžném prohlédnutí již zpracovaných dat bylo zjištěno, že výše uvedená funkce na kontrolu validity nefunguje zcela správně, jelikož neodstranila záznamy které měly ve sloupečku pro geometrii uvedenou hodnotu *NULL*. Bylo tedy nutné provést dodatečné pročištění dat níže uvedeným příkazem:

```
DELETE FROM adresni_mista WHERE geom IS NULL
```

Správní obvody Prahy

Data byla stažena ze schématu *ruian_praha* (vrstva *spravniobvody*) do nově vytvořené tabulky *obvody*. Geometrie dat je *multipolygon* a představují jednotlivé pražské správní obvody. Z původních dat byly zkopírovaný pouze sloupečky *ogc_fid*, *název* a sloupeček s geometrií *geom*.

```
CREATE TABLE obvody AS
    SELECT ogc_fid, nazev, geom
    FROM ruian_praha.spravniobvody
```

Následně byl nad sloupečkem *ogc_fid* nastaven primární klíč.

```
ALTER TABLE obvody ADD PRIMARY KEY(ogc_fid)
```

Na závěr byla provedena kontrola validity geometrie, která vyšla negativní.

```
SELECT nazev FROM obvody WHERE NOT st_isvalid(geom)
```

3.2 Otevřená data

Všechna data z tohoto zdroje byla stažena ve formátu geoJSON a byla převážně v systému WGS84 (výjimka: vrstva *metro* byla od počátku v S-JTSK). Postup při zpracování těchto dat byl obdobný. Data byla importována do příslušného schématu pomocí nástroje *ogr2ogr* a přímo v databázi byly v tabulkách promazány nepotřebné sloupečky a

přejmenován sloupeček s geometrií, který po nahrání dávkou získal uživatelsky nepřívětivý název *wkb_geometry*.

Během importu do databáze bylo u některých dat nutné provést transformaci ze systému WGS84 (EPSG: 4326) do S-JTSK (EPSG: 5514). Transformaci zajišťovala část kódu *-t_srs 'EPSG:5514'*. Nebyla-li transformace nutná, nahradila se tato část kódu za *-a_srs 'EPSG:5514'*. Při importu byl automaticky vytvořen sloupeček s primárním klíčem *ogc_fid* a byly vytvořeny prostorové indexy. Níže je uveden obecný kód pro nahrání souboru do příslušného schématu databáze (s transformací do S-JTSK):

```
ogr2ogr -f "PostgreSQL" PG:"dbname=pgis_uzpd_user=uzpd18_d_host=geo102.fsv.
    cvut.cz" -t_srs 'EPSG:5514' "input.json" -nln uzpd18_d.table
```

Jelikož geometrie všech vrstev byla *bod*, pro kontrolu validity byla pro všechny vrstvy použita stejná funkce:

```
SELECT ogc_fic FROM nazev_tabulky WHERE NOT st_isvalid(geom)
```

Dětská hřiště Praha

Příkaz pro smazání sloupečků, které pro další práci s daty nebyly nutné:

```
ALTER TABLE detska_hriste
    DROP COLUMN url,
    DROP COLUMN name,
    DROP COLUMN perex,
    DROP COLUMN content,
    DROP COLUMN address,
    DROP COLUMN properties,
    DROP COLUMN image
```

Přejmenování sloupečku s geometrií:

```
ALTER TABLE detska_hriste
    RENAME wkb_geometry TO geom
```

Výsledná tabulka má tyto sloupce:

detska_hriste			
ogc_fid	geom	id	district

Lékárny a zdravotnická zařízení v Praze

Příkaz pro smazání sloupečků, které pro další práci s daty nebyly nutné:

```
ALTER TABLE zdrav_zarizeni
    DROP COLUMN id,
    DROP COLUMN address,
    DROP COLUMN email,
    DROP COLUMN web,
    DROP COLUMN telephone,
    DROP COLUMN opening_hours
```

Přejmenování sloupečku s geometrií:

```
ALTER TABLE zdrav_zarizeni
    RENAME wkb_geometry TO geom
```

Výsledná tabulka má tyto sloupce:

zdrav_zarizeni				
ogc_fid	geom	name	type	district

Vstupy do metra

Jelikož stažená data byla již v S-JTSK, část kódu zajišťující transformaci při importu byla nahrazena za *-a_srs 'EPSG:5514'*. Příkaz pro smazání sloupečků, které pro další práci s daty nebyly nutné:

```
ALTER TABLE metro
    DROP COLUMN objectid,
    DROP COLUMN vstupy_kod,
    DROP COLUMN vstupy_vest_kod,
    DROP COLUMN vstupy_vest_nazev,
    DROP COLUMN vstupy_vazba_bus,
    DROP COLUMN vstupy_vazba_csad,
    DROP COLUMN vstupy_vazba_kr,
    DROP COLUMN vstupy_vazba_pr,
    DROP COLUMN vstupy_vazba_privoz,
    DROP COLUMN vstupy_vazba_taxi,
    DROP COLUMN vstupy_vazba_tram,
    DROP COLUMN vstupy_vazba_vlak,
    DROP COLUMN zast_uzel_cislo,
    DROP COLUMN vstupy_popis,
    DROP COLUMN poskyt,
    DROP COLUMN vstupy_mimo_provoz
```

Přejmenování sloupečku s geometrií a názvu stanic:

```
ALTER TABLE metro
    RENAME wkb_geometry TO geom,
    RENAME vstupy_uzel_nazev TO nazev,
    RENAME vstupy_linka TO linka
```

Výsledná tabulka má tyto sloupce:

metro			
ogc_fid	geom	nazev	linka

3.3 Geoportál Praha

Data získaná z tohoto portálu byla stažena ve formátu geoJSON v systému JTSK. Zpracování dat probíhalo obdobně jako u předchozího zdroje dat s výjimkou, že při importu nebylo nutné provádět transformaci. Obecný kód pro nahrání souboru do příslušného schématu databáze (bez transformace do S-JTSK):

```
ogr2ogr -f "PostgreSQL" PG:"dbname=pgis_uzpd user=uzpd18_d host=geo102.fsv.
    cvut.cz" -a_srs 'EPSG:5514' "input.json" -nln uzpd18_d.table
```

Mapa košů na tříděný odpad

Příkaz pro smazání sloupečků, které pro další práci s daty nebyly nutné:

```
ALTER TABLE odpad
    DROP COLUMN objectid,
    DROP COLUMN id,
    DROP COLUMN stationnumber,
    DROP COLUMN stationname,
    DROP COLUMN citydistrictruiancode
```

Přejmenování sloupečku s geometrií:

```
ALTER TABLE odpad
    RENAME wkb_geometry TO geom
```

Na závěr byla provedena kontrola validity geometrie.

```
SELECT ogc_fic FROM nazev_tabulky WHERE NOT st_isvalid(geom)
```

Výsledná tabulka má tyto sloupce:

odpad			
ogc_fid	geom	citydistrict	pristup

Záplavové území 2013

Příkaz pro smazání sloupečků, které pro další práci s daty nebyly nutné:

```
ALTER TABLE zaplava2013
    DROP COLUMN objectid,
    DROP COLUMN nazev,
    DROP COLUMN typ
```

Přejmenování sloupečku s geometrií:

```
ALTER TABLE zaplava2013
    RENAME wkb_geometry TO geom
```

Opět bylo nutné zkontrolovat validitu geometrie (polygonů). Níže uvedeným příkazem byly zjištěny chyby v datech:

```
SELECT id, geom, st_isvalidreason(geom) FROM uzpd18_d.zaplava2013 WHERE NOT
    st_isvalid(geom)
```

Ukázalo se, že nevalidní byly 4 polygony z celkových 86. Jako příčina těchto chyb se ukázala tzv. *Ring-self Intersection* polygonů. Nevalidní polygony byly opraveny vytvořením *bufferu* o velikosti 0.

```
UPDATE uzpd18_d.zaplava2013 SET geom = st_buffer(geom, 0.0) WHERE NOT
    st_isvalid(geom)
```

Výsledná tabulka má tyto sloupce:

zaplava2013			
ogc_fid	geom	shape_length	shape_area

4 SQL dotazy

1) Kolik adresních míst bylo zatopeno během povodní v roce 2013?

```
SELECT COUNT(a.kod)
FROM uzpd18_d.adresni_mista AS a
JOIN (SELECT * FROM uzpd18_d.zaplava2013) AS z
ON st_intersects(z.geom, a.geom);
```

1042

2) Jaká zdravotnická zařízení byla zasažena povodněmi v roce 2013? Vypište jejich název a správní obvod, ve kterém se nacházejí.

```
SELECT zz.name, zz.district
FROM uzpd18_d.zdrav_zarizeni AS zz
JOIN uzpd18_d.zaplava2013 AS za
ON st_intersects(zz.geom,za.geom);
```

Dr.Max LÉKÁRNA, praha-8

3) Jaké procento území Prahy zaujímá záplavové území z roku 2013?

```
SELECT ROUND(
(
SELECT sum(st_area(geom)) FROM uzpd18_d.zaplava2013
)::numeric / (
SELECT sum(st_area(geom)) FROM uzpd18_d.obvody
)::numeric, 2)*100 AS procento_zaplava;
```

5%

4) Kolik košů na tříděný odpad se nachází na území Prahy 1?

```
SELECT COUNT(o.ogc_fid)
FROM uzpd18_d.odpad AS o
JOIN (SELECT * FROM uzpd18_d.obvody WHERE nazev = 'Praha 1') AS p
ON st_intersects(p.geom, o.geom);
```

821

5) Kolik adresních míst leží ve vzdálenosti do 500 m od výlezu ze stanice metra Depo Hostivař?

```
SELECT count(distinct a.kod)
FROM uzpd18_d.adresni_mista AS a
JOIN uzpd18_d.metro AS m
ON st_dwithin(a.geom, m.geom, 500)
WHERE m.nazev = 'Depo Hostivař';
```

48

6) Jaká stanice metra na Praze 4 má nejvíce vstupů?

```
SELECT m.nazev FROM uzpd18_d.metro AS m
JOIN (SELECT * FROM uzpd18_d.obvody
WHERE nazev = 'Praha 4') AS p
ON st_intersects(p.geom, m.geom)
GROUP BY m.nazev
ORDER BY COUNT(m.ogc_fid) DESC
LIMIT 1;
```

Budějovická

7) Která zdravotnická zařízení leží do 10 m od stanic metra A? V které městské části se tato zařízení nacházejí?

```
SELECT z.name, z.district
FROM uzpd18_d.zdrav_zarizeni AS z
JOIN uzpd18_d.metro AS m
ON st_dwithin(z.geom, m.geom, 10)
WHERE m.linka LIKE '%A%';
```

Dr.Max LÉKÁRNA, praha-1
BENU Lékárna OC Atrium Flora, praha-3

8) Kolik dětských hřišť se nachází ve vzdálenosti do 2 kilometrů od nejbližší nemocnice?

```
WITH
    n AS (
        SELECT * FROM uzpd18_d.zdrav_zarizeni AS z
        WHERE type LIKE '%nemocnice%')
SELECT COUNT(DISTINCT(h.ogc_fid))
FROM uzpd18_d.detska_hriste AS h, n
WHERE n.geom && st_expand(h.geom, 2000);
```

61

9) Kolik adresních míst má lékárnu a dětské hřiště do vzdálenosti 500 m?

```
WITH
    buff_h AS (
        SELECT st_buffer(geom, 500) AS geom
        FROM uzpd18_d.detska_hriste AS h),
    buff_z AS (
        SELECT st_buffer(geom, 500) AS geom
        FROM uzpd18_d.zdrav_zarizeni AS z
        WHERE type = 'Lékárna')
SELECT COUNT(DISTINCT(kod)) FROM uzpd18_d.adresni_mista AS a
INNER JOIN buff_h ON st_intersects(a.geom, buff_h.geom)
INNER JOIN buff_z ON st_intersects(a.geom, buff_z.geom);
```

25130

10) Který pražský správní obvod má nejlepší poměr počtu košů na tříděný odpad vzhledem ke své rozloze? Uveďte název obvodu, jeho rozlohu, počet košů a poměr těchto dvou hodnot.

```
WITH
    oo AS (
        SELECT ob.nazev, count(od.ogc_fid), ob.geom FROM obvody AS ob
        JOIN uzpd18_d.odpad AS od
        ON st_intersects(ob.geom, od.geom)
        GROUP BY ob.nazev, ob.geom
        ORDER BY COUNT(od.ogc_fid) DESC),
    area AS (
        SELECT obvody.nazev, st_area(obvody.geom), geom from obvody)
SELECT oo.nazev, ROUND(area.st_area), count, ROUND(area.st_area/count) AS
    ratio
FROM oo
JOIN area ON oo.nazev = area.nazev
GROUP BY oo.nazev, area.st_area, count, area.st_area/count
ORDER BY area.st_area/count DESC
LIMIT 1;
```

Praha 22, 33660132, 70, 480859

11) Na území kterého pražského správního obvodu se nachází největší množství dětských hřišť a kolik to je?

```
SELECT o.nazev, COUNT(distinct dh.ogc_fid)
FROM uzpd18_d.obvody AS o
JOIN uzpd18_d.detska_hriste AS dh
ON st_intersects(dh.geom, o.geom)
GROUP BY o.nazev
HAVING COUNT(distinct dh.ogc_fid) = (
    SELECT COUNT(distinct dh.ogc_fid) FROM uzpd18_d.obvody AS o
    JOIN uzpd18_d.detska_hriste AS dh
    ON st_intersects(dh.geom, o.geom)
    GROUP BY o.nazev
    ORDER BY COUNT(distinct dh.ogc_fid) DESC
    LIMIT 1);
```

Praha 3, 14
Praha 4, 14

12) Jaké je id nejbližšího dětského hřiště pro adresní místo s kódem 22560840? V jaké leží vzdálenosti?

```
SELECT dh.id, ROUND(st_distance(a.geom, dh.geom)) AS distance
FROM uzpd18_d.detska_hriste AS dh, uzpd18_d.adresni_mista AS a
WHERE a.kod = '22560840' AND st_distance(a.geom, dh.geom) = (
    SELECT st_distance(a.geom, dh. geom)
    FROM detska_hriste AS dh, uzpd18_d.adresni_mista AS a
    WHERE a.kod = '22560840'
    ORDER BY st_distance(a.geom, dh.geom)
    LIMIT 1);
```

120, 506 m

5 Závěr

Na základě volně dostupných zdrojů dat byly navrženy a zpracovány tematické vrstvy. Nad těmito vrstvami byly následně aplikovány prostorové a atributové dotazy. Semestrální projekt nám přinesl další zkušenosti při zpracování prostorových dat a jejich uchování v databázích.

Během zpracování se vyskytlo několik problémů. Prvním problémem se ukázalo býti zvolené téma. Samo o sobě je velmi široké a při jeho volbě nás nenapadlo, že by v tom mohla být potíž. Pro naši analýzu bylo k dispozici ke stažení velké množství dat, z nichž jsme se rozhodly použít jen několik zdrojů a ty zpracovat podrobněji. Použití velkého množství dat (a že se nám jich líbilo hodně) by zapříčinilo, že výsledek naší práce by nebyl tak kvalitně zpracován, jak bychom si přály. Nakonec se ukázalo, že použité datové vrstvy neobsahovaly takové množství užitečných informací, jak jsme předpokládaly, a bylo nutné smazat většinu sloupečků v tabulkách. Výsledek projektu odpovídá zvolenému tématu jen částečně, jelikož se nám povedla pouze analýza bydlení v Praze na základě zvolených tématických vrstev, nikoli komplexní jak jsme původně zamýšlely.

Jako zádrhel při zpracování se ukázal import dat ve formátu geoJSON. Nikde jsme neobjevily přesný návod, jak importovat data tohoto typu do postGIS databáze přes příkazovou řádku, je-li součástí dávky i přihlašovací jméno a heslo do databáze. Správný zápis příkazu se nakonec podařilo nalézt.

6 Přílohy

- Příloha č. 1: Prezentace (prezentace.pdf)
- Příloha č. 2: SQL dávka (davka.sql)

7 Zdroje

1. *Otevřená data* [online] [cit. 28. 1. 2019].
   Dostupné z: http://www.geoportalpraha.cz/

2. *Datové sady – Národní katalog otevřených dat (NKOD)* [online] [cit. 28. 1. 2019].
   Dostupné z: https://data.gov.cz/

3. *Školení postGIS pro začátečníky* [online] [cit. 30. 1. 2019].
   Dostupné z: http://training.gismentors.eu/

4. *Školení postGIS pro pokročilé* [online] [cit. 30. 1. 2019].
   Dostupné z: http://training.gismentors.eu/

5. *155UZPD / Semestrální projekt* [online] [cit. 30. 1. 2019].
   Dostupné z: http://geo.fsv.cvut.cz/

6. *LaTeX/Source Code Listing* [online] [cit. 28. 1. 2019].
   Dostupné z: https://en.wikibooks.org/