

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE  
FAKULTA STAVEBNÍ  
obor Geoinformatika



## DOKUMENTÁCIA

ÚVOD DO ZPRACOVÁNÍ PROSTOROVÝCH DAT  
Semestrálny projekt

Skupina C  
Markéta Sedláčková  
Michal Bečička  
Michal Bodnár  
Jiří Kratochvíl  
Filip Zavadil

# Obsah

1 Úvod .....	2
1.1 Cieľ projektu .....	2
1.2 Zdroj dát a práca s nimi .....	2
2 Tvorba tématických vrstiev .....	3
2.1 Vrstva c10.gastronomy .....	3
2.2 Vrstva c10.counties .....	5
3 Kontrola konzistencie dát .....	6
2.1 Vrstva c10.gastronomy .....	6
2.2 Vrstva c10.counties .....	7
4 SQL dotazy .....	9
5 Záver .....	10
6 Zdroje .....	11

# 1 ÚVOD

## 1.1 Cieľ projektu

Cieľom seminárneho projektu predmetu Úvod do zpracování prostorových dat (ďalej len UZPD) bolo:

- Navrhňte a vytvorte tématické vrstvy (např. vodní toky, vodní plochy, lesy, silnice, železnice a pod.) na základě dat OSM (viz [cvičná databáze](#) pgis\_student schéma [osm](#)). Pro tento účel byla na serveru 'josef' založena databáze pgis\_osm ([poznámky](#)).
- Aplikujte testy datové integrity a odstraňte případné nekonzistence v datech.
- Vytvorte tutoriál pro výuku [PostGIS](#) - tj. sadu atributových a prostorových dotazů nad databází pgis\_osm.

## 1.2 Zdroj dat a práca s nimi

Zdrojom dát bola cvičná databáza `pgis_osm` naplnená datami OpenStreetMaps<sup>1</sup> (ďalej len OSM) umiestnená na serveri `josef.fsv.cvut.cz`. Práca prebiehala prostredníctvom jazyka SQL v Postgise<sup>2</sup>, relačno-objektovom rozšírení databázového serveru PostgreSQL<sup>3</sup> o geometrické dátové členy. Prístup k databázi prebiehal pomocou `psql`<sup>4</sup> klientskej interaktívnej aplikácie prostredníctvom terminálu.

Práca v databázi prebiehala v schémate `c10`, pod ktorým boli pre jednotlivých členov skupiny `c10` pridelené práva pre tvorbu tabuliek (`CREATE`), ich modifikáciu (`ALTER`, `UPDATE`), mazanie dát (`DELETE`), prípadne celej tabuľky (`DROP`).

---

<sup>1</sup><http://www.openstreetmaps.org>

<sup>2</sup><http://postgis.refraction.net>

<sup>3</sup><http://www.postgresql.org>

<sup>4</sup> <http://www.postgresql.org/docs/8.4/static/app-psql.html>

## 2 Tvorba tématických vrstiev

Boli vytvorené 2 tématické vrstvy : c10.gastronomy  
c10.counties

Plnenie tabuliek datami z OSM bolo definované pomocou kľúčového slova *key* a jeho hodnoty *value*.

### 2.1 vrstva c10.gastronomy

- vytvorenie tabuľky naplnením dát z OSM

```
CREATE TABLE c10.gastronomy
AS SELECT osm_id,name,amenity,way
FROM czech_point
WHERE amenity IN
('restaurant','cafe','fast_food','pub','bar','biergarten');
```

- vytvorenie primárneho kľúča osm\_id

```
ALTER TABLE c10.gastronomy
ADD PRIMARY KEY (osm_id);
```

- aktualizácia tabuľky Populate\_geometry\_columns

```
SELECT Populate_geometry_columns('c10.gastronomy'::regclass);
```

- vytvorenie indexu nad atribútom geometrie way pre urýchlenie procesu priestorového vyhľadávania:

```
CREATE INDEX gastronomy_gist
ON c10.gastronomy
USING GIST(way)
```

- pridelenie práv pre ostatných členov skupiny k modifikácii tabuľky

```
GRANT ALL ON c10.gastronomy TO bodnami1, sedlama8,kratoj25,zavadfil;
```

- premenovanie atribútu amenity na kind

```
ALTER TABLE c10.gastronomy
RENAME amenity TO kind
```

- premenovanie hodnoty atribútu kind biergarten na beergarden v zmysle dodržiavať konvenciu v anglickom jazyku

```
UPDATE c10.gastronomy
SET kind = 'beergaden'
WHERE kind = 'biergarten'
```

Vizualizáciou v <sup>1</sup>QGise:



---

<sup>1</sup><http://www.qgis.org>

## 2.2 vrstva c10. counties

- vytvorenie tabuľky naplnením dát z OSM

```
CREATE TABLE c10.counties
AS SELECT osm_id,name,way
FROM czech_polygon
WHERE place = 'county'
```

- vytvorenie primárneho kľúča osm\_id

```
ALTER TABLE c10.counties
ADD PRIMARY KEY (osm_id);
```

- aktualizácia tabuľky Populate\_geometry\_columns

```
SELECT Populate_geometry_columns('c10.counties'::regclass);
```

- vytvorenie indexu nad atribútom geometrie way pre urýchlenie procesu priestorového vyhľadávania:

```
CREATE INDEX counties_gist
ON c10.gastronomy
USING GIST(way)
```

- pridelenie práv pre ostatných členov skupiny k modifikácii tabuľky

```
GRANT ALL ON c10.counties TO bodnam1, sedlama8,kratoj25,zavadfil;
```

Vizualizáciou v QGise :



## 3 Kontrola konzistencie dát

### 3.1 Vrstva c10.gastronomy

#### 1. odstránenie prvkov, u ktorých atribút name má hodnotu NULL

```
DELETE FROM c10.gastronomy
WHERE name IS NULL
```

#### 2. snaha o jednotnosť atribútu name

Jednotlivé názvy pohostinství sa líšili v tom, že zatiaľ čo nejaké sa začínali na malé písmeno, tak ostatné mali začiatkové písmeno veľké. Takisto konvencia veľkých resp. malých písmen v prípade za predložkou ako začiatkom názvu bola rozdielna. Preto bola snaha zjednotiť jednotlivé názvy pre ľahšiu orientáciu užívateľa.

Došlo k upraveniu atribútu name na veľké počiatočné písmeno a zvyšné písmená malé :

```
UPDATE c10.gastronomy
SET name = upper(substring(name from 1 for 1)) ||
lower(substring(name from 2 for length(name)))
```

```
UPDATE 1452
```

Špeciálnym prípadom bol názov name = 'McDonalds', u ktorého sa nachádzali rôzne modifikácie tohto názvu, preto došlo k jednotnej úprave:

```
UPDATE c10.gastronomy
SET name='McDonalds'
WHERE name LIKE 'Mc%'
```

## 3.2 Vrstva c10.counties

Po vytvorení tabuľky pomocou dat OSM bolo zistené, že tabuľka neobsahuje žiadne data, je prázdna.

```
SELECT name
FROM c10.counties
```

```
name
-----
(0 rows)
```

Táto situácia sa určite neočakávala, nakoľko sa čakalo, že polygóny krajov ČR budú v OSM definované. Nakoniec sa nám s pomocou Martina Landy podarilo zistiť, že vytvorenie vrstvy stroskotalo na probléme relácií dát definujúcich jednotlivé administratívne členy. Ďalším faktom bolo to, že tag `place` sa už v súčasnosti nepoužíva, pre definíciu administratívnych celkov sa používa kľúčové slovo `administrative` v kombinácii s hodnotou atribútu `admin_level`.

Riešením problému bol nový dátový zdroj – databáza `fgcz` naplnené datami datasetu FreeGeodataCZ. Martinom Landom bol prevedený import tabuliek `kraje` a `okresy` do schématu `public` databáze `pgis_osm`. Následne lokálnou kópiou tabuľky `kraje` bola vytvorená nová tabuľka `c10.counties`:

```
DROP TABLE c10.counties

CREATE TABLE c10.counties
AS SELECT *
FROM KRAJE
```

Vizualizáciou oboch pracovných vrstiev v <sup>5</sup>QGise sa odhalil ďalší problém – zatiaľ čo vrstva `c10.gastronomy` bola situovaná v referenčnom systéme *Google Projection* (`srid = 900913`), tak data vrstvy `c10.counties` boli v referenčnom systéme *WGS-84* (`srid = 4326`). Bolo teda potrebné vrstvu `c10.counties` pretransformovať do referenčného systému *Google Projection*.

```
DROP TABLE c10.counties

CREATE TABLE c10.counties
SELECT *,transform(wkb_geometry,900913)
FROM kraje
```

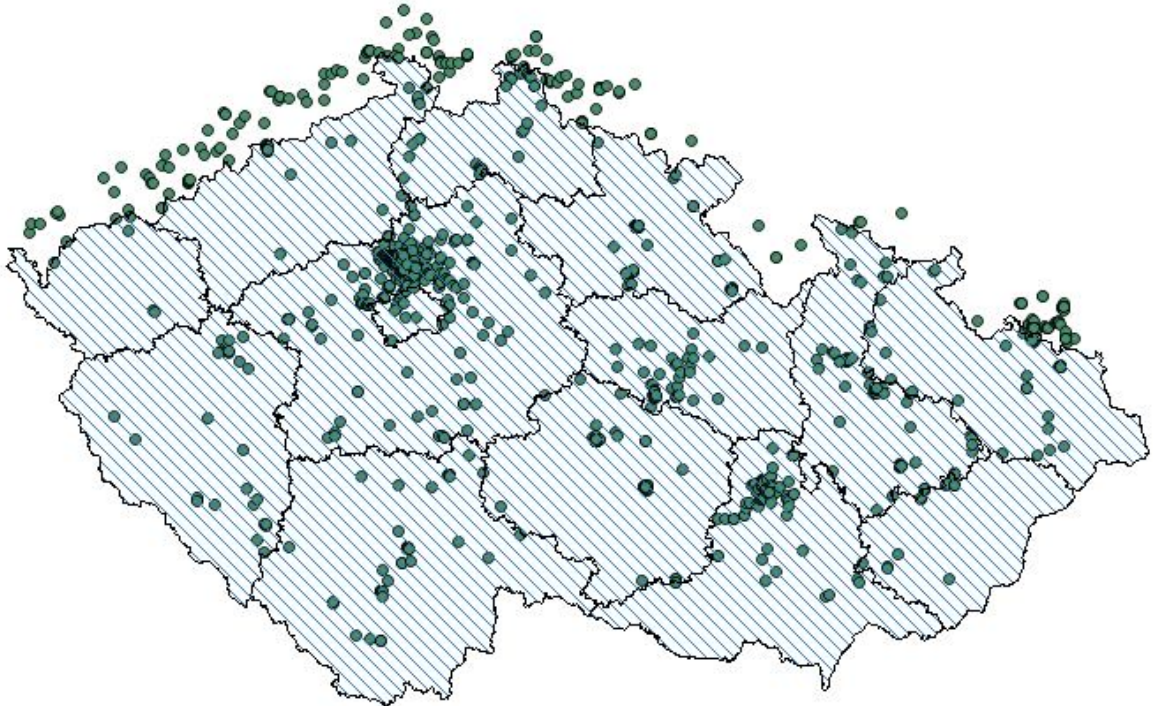
Opätovnou vizualizáciou oboch vrstiev v QGise sa vyskytol ďalší problém. Vrstva `c10.gastronomy` bola oproti vrstve `c10.counties` posunutá v smere na sever asi o 2 km. Dôvod tejto chyby sa nám však po dobu trvania projektu napriek veľkej snahe zistiť nepodaril. Po návrhu Martina Landy, že chyba by sa mohla vyskytnúť v tom, že data vrstvy `c10.counties` boli transformované z S-JTSK do WGS-84 bez geodetického dátumu, nám boal poskytnutá opätovne nová vrstva `kraje1` opravená o spomínaný problém.

```
DROP TABLE c10.counties

CREATE TABLE c10.counties
SELECT *,transform(wkb_geometry,900913)
FROM kraje1
```



Následnou vizualizáciou v QGise sa však ukázalo, že príčinu chyby sa nepodarilo odhaliť.



Tento problém sa nám bohužiaľ do doby odovzdania projektu napriek veľkej snahe nepodarilo vyriešiť.

## 4 SQL dotazy

Vzhľadom k nevyriešenému problému s vrstvou `c10.counties` a priestorovou polohou jej dát tvorba priestorových dotazov bola obmedzená.

### 1. Koľko je v ČR pohostinství 'McDonalds'.

```
SELECT count(*)
FROM c10.gastronomy
WHERE name='McDonald';
-- 29
```

### 2. Koľko reštaurácií začína na predložku 'U'?

```
SELECT count(*)
FROM c10.gastronomy
WHERE kind='restaurant'
AND name LIKE 'U %';
-- 255
```

### 3. Zoraď podľa abecedy názvy reštaurácií.

```
SELECT name
FROM c10.gastronomy
ORDER BY name;
```

### 4. Aké su druhy gastronomických zariadení?

```
SELECT DISTINCT kind
FROM c10.gastronomy
-- restaurant, fast_food, cafe, beergarden
```

### 5. Ktorý kraj má najviac okresov?

```
SELECT nazkr, count(nazok)
FROM okresy
GROUP BY nazkr
ORDER BY count(nazok)
-- Stredocesky 12
```

### 6. Aká je priemerná rozloha jedného okresu v ČR?

```
SELECT ROUND(AVG(vymera)) FROM okresy;
-- 102424
```

### 7. Aká je priemerná rozloha obce v Ústeckom kraji?

```
SELECT ROUND(AVG(vymera)) FROM okresy WHERE nazkr_a='Ustecky';
-- 76213
```

### 8. Koľko je v Prahe kavární?

```
SELECT COUNT(*) FROM c10.gastronomy AS gastronomy
JOIN c10.counties AS okresy
ON ST_within(gastronomy.way, okresy.wkb_geometry)
WHERE kind='cafe';
```

## 5 Záver

Cieľ tohto projektu : vytvorenie tématických vrstiev, kontrola konzistencie dát a tvorba sql dotazy sa podaril splniť. Problém, ktorý sa naskytl, a to zlé nareferencovanie vrstvy `c10.counties` do systému *Google Projection* spôsobil značné obmedzenie v tvorbe priestorových dotazov. V budúcnosti je naším cieľom tento problém vyriešiť.

## 6 Zdroje

[1] domáce stránky predmetu UZPD  
<http://gama.fsv.cvut.cz/wiki/index.php/153UZPD>

[2] zdroj hľadania problémov v PostGise  
<http://www.wildsong.biz/index.php?title=PostGIS>

[3] wikipedia  
<http://www.wikipedia.org/>