

Úvod do zpracování prostorových dat

Semestrální projekt

# DOKUMENTACE

Kateřina Hynková

Martin Lžíčář

Tomáš Tichý

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta stavební

Katedra mapování a kartografie

11. 6. 2011

# Obsah

<b>1</b>	<b>Úvod</b>	<b>3</b>
1.1	Zadání . . . . .	3
<b>2</b>	<b>Zdroje dat</b>	<b>3</b>
2.1	Schéma osm . . . . .	3
2.2	Schéma gis1 . . . . .	3
<b>3</b>	<b>Tvorba tematických vrstev a jejich validace</b>	<b>4</b>
3.1	Bodové vrstvy . . . . .	4
3.1.1	Hrady a zámky . . . . .	4
3.1.2	Koupání . . . . .	7
3.1.3	Stravování . . . . .	8
3.2	Liniové vrstvy . . . . .	8
3.2.1	Silnice . . . . .	8
3.3	Polygonové vrstvy . . . . .	9
3.3.1	Obce . . . . .	9
<b>4</b>	<b>Dotazy</b>	<b>9</b>
4.1	Atributové dotazy . . . . .	9
4.2	Prostorové dotazy . . . . .	10
<b>5</b>	<b>Použitý software</b>	<b>21</b>
<b>6</b>	<b>Závěr</b>	<b>21</b>
<b>7</b>	<b>Seznam příloh</b>	<b>21</b>

# 1 Úvod

Tato dokumentace si klade za cíl seznámit čtenáře s průběhem vzniku semestrálního projektu pro předmět Úvod do zpracování prostorových dat (153UZPD). Tento předmět je vypisován Katedrou mapování a kartografie pro studenty 3. ročníku oboru Geoinformatika na Fakultě stavební Českého vysokého učení technického v Praze.

Z poskytnutých dat byly vytvořeny tematické vrstvy a po kontrole jejich validity nad nimi byly provedeny atributové a prostorové dotazy.

## 1.1 Zadání

- Navrhněte a vytvořte tematické vrstvy na základě dat OpenStreetMap (OSM)
- Aplikujte testy datové integrity a odstraňte případné nekonzistence v datech
- Vytvořte tutoriál pro výuku PostGIS - tj. sadu atributových a prostorových dotazů nad databází `pgis_uzpd`

## 2 Zdroje dat

Data byla získána z cvičné databáze PostGIS `pgis_student`<sup>1</sup> umístěné na serveru `geo.102`<sup>2</sup>.

### 2.1 Schéma osm

Schéma `osm` obsahuje data z projektu OpenStreetMap<sup>3</sup> pro území České republiky ze dne 9. 3. 2011. OpenStreetMap používá souřadnicový systém Google Mercator (SRID 900913).

### 2.2 Schéma gis1

Schéma `gis1` obsahuje data ze cvičení k předmětu GIS1 (153GIS1). Tato data jsou v souřadnicovém systému S-JTSK Krovak East North (SRID 102067).

---

<sup>1</sup>[http://geo101.fsv.cvut.cz/gwiki/Cvičná\\_databáze\\_PostGIS#gis1](http://geo101.fsv.cvut.cz/gwiki/Cvičná_databáze_PostGIS#gis1)

<sup>2</sup><http://geo102.fsv.cvut.cz/>

<sup>3</sup>[www.openstreetmap.org](http://www.openstreetmap.org)

## 3 Tvorba tematických vrstev a jejich validace

Pro provádění analýz jsme se rozhodli vytvořit 5 tematických vrstev. Volbu jejich obsahu jsme podmínili určitou praktickou využitelností. Vrstvy jsme koncipovali tak, aby je ve výsledku bylo možné použít k vyhledávání turistických cílů (hradů a zámků) a dále koupališť či stravovacích zařízení v jejich okolí.

Pro zjednodušení příkazů a dotazů byla nastavena cesta ke schémátům:

```
SET SEARCH_PATH TO e11, public, osm
```

Pro každou vytvořenou tematickou vrstvu byl definován primární klíč, vytvořen prostorový index a aktualizována metadatová tabulka `geometry_columns`.

Všechny konkrétní příkazy a dotazy, které vedly k vytvoření validních tematických vrstev, se nachází v příloženém SQL dávkovém souboru.

### 3.1 Bodové vrstvy

#### 3.1.1 Hrady a zámky

Do vrstvy hradů a zámků byly vybrány ze schématu `osm` prvky s atributem `castle` (hrady a zámky) ze sekce `historic` (historická místa). Tyto atributy mají jak polygonové, tak bodové prvky, proto byly v první fázi vytvořeny 2 vrstvy:

- `castle_p` (polygonová vrstva)

```
CREATE TABLE e11.castle_p
AS SELECT osm_id, way, name
FROM czech_polygon
WHERE historic='castle'
```

- `castle_b` (bodová vrstva)

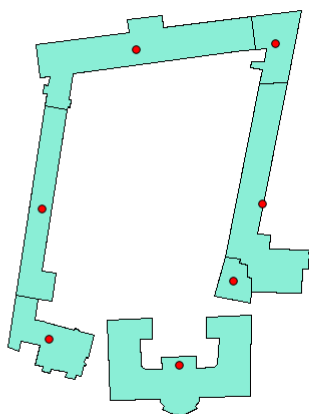
```
CREATE TABLE e11.castle_b
AS SELECT osm_id, way, name
FROM czech_point
WHERE historic='castle'
```

Abychom mohli provést kontroly, zda se ve výsledné vrstvě nenachází pro 1 objekt 2 body, odstranili jsme z nich prvky, jejichž atribut `name` měl hodnotu `NULL`.

Pro účely zamýšlených analýz plně dostačují data ve formě bodů. Polygonová vrstva `castle_p` byla tedy pomocí funkce `ST_CENTROID`, která vrací těžiště prvku,

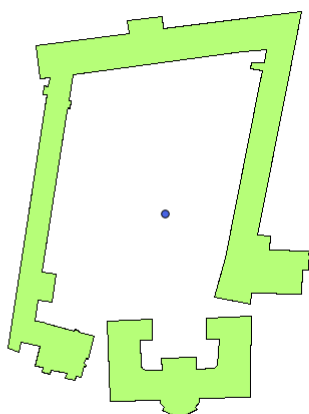
převědena na vrstvu bodovou.

Některé objekty ovšem tvoří více polygonů a v tomto případě funkce `ST_CENTROID` vrací těžiště jednotlivých prvků.



*Obr.1: Aplikace funkce `ST_Centroid` pro objekt tvořený více polygony*

Z tohoto důvodu jsme tyto polygony nejprve spojili funkcí `ST_UNION` a teprve poté použili funkci `ST_CENTROID`.



*Obr.2: Aplikace funkce `ST_Centroid` na sloučený polygon*

Body získané tímto způsobem jsme vložili do bodové vrstvy hradů a zámků hrady\_zamky. Prvkům, které vznikly sjednocením více polygonů, bylo vygenerováno id.

- Vložení bodů vzniklých z 1 polygonu

```
INSERT INTO hrady_zamky
SELECT osm_id, ST_CENTROID(way) AS way, name
FROM castle_p
WHERE name NOT IN ('Hrad Střekov',
'Vargač',
'zámek Tloskov',
'Bučovický zámek',
'Zámek Hořín',
'Nové Hrady',
'Dačice',
'Zámek Veltrusy',
'Český Šternberk',
'Nový zámek',
'Bítov',
'Zámek Benešov nad Ploučnicí',
'Hrad Kunětická Hora',
'Zámek Dobříš',
'zámek Velké Meziříčí')
```

- Vložení bodů vzniklých z více polygonů

```
CREATE TABLE docasna
AS SELECT ST_CENTROID(ST_UNION(way)) AS way, name
FROM castle_p
WHERE name IN ('Hrad Střekov',
'Vargač',
'zámek Tloskov',
'Bučovický zámek',
'Zámek Hořín',
'Nové Hrady',
'Dačice',
'Zámek Veltrusy',
'Český Šternberk',
'Nový zámek',
'Bítov',
'Zámek Benešov nad Ploučnicí',
'Hrad Kunětická Hora',
'Zámek Dobříš',
```

```
'zámek Velké Meziříčí')
GROUP BY name

ALTER TABLE docasna ADD COLUMN gid serial;

INSERT INTO hrady_zamky
SELECT gid, way, name
FROM docasna
```

Poté jsme provedli kontrolu, zda se vrstvě nenachází pro jeden objekt 2 body. Nejprve jsme vyhledali body se stejným atributem `name`.

```
SELECT name, COUNT(name)
FROM hrady_zamky
GROUP BY name
ORDER BY count(name)
```

Dále jsme vybrali body z původní bodové vrstvy `castle_b`, které leží uvnitř polygonů původní polygonové vrstvy `castle_p` a vizuálně zkontrolovali, zda pro jeden objekt neexistují 2 body s různým atributem `name`.

```
CREATE TABLE kontrola
AS SELECT ST_INTERSECTION(h.way,c.way)
FROM castle_b AS h
JOIN castle_p AS c
ON ST_INTERSECTS(h.way,c.way)
```

Pokud byly ve vrstvě `hrady_zamky` nalezeny 2 body definující 1 objekt, byl jeden z bodů odstraněn.

### 3.1.2 Koupání

Do vrstvy koupališť byly ze schématu `osm` vloženy prvky s atributem `leisure` (volný čas) o hodnotě `water_park` (aquapark) nebo `swimming_pool` (koupaliště).

Tyto objekty se v databázi stejně jako hrady a zámky objevují ve formě bodů i polygonů, proto jsme je stejným postupem jako v předchozím případě převedli na výhradně bodové prvky.

Ověření, zda se ve vrstvě nenachází 2 body definující 1 objekt, bylo provedeno na základě stejných názvů. Většina prvků v databázi ovšem nemá vyplněn atribut `name`, proto nebyla kontrola úplná. Vrstva je přesto validní.

### 3.1.3 Stravování

Do vrstvy stravovacích zařízení byly ze schématu `osm` vloženy prvky s atributem `amenity` (občanská vybavenost) o hodnotě `restaurant` (restaurace), `food_court` (místo s více stravovacími zařízeními), `fast_food` (fast food), `pub` (hospoda) nebo `biergarten` (venkovní posezení).

Tyto objekty se v databázi stejně jako hrady a zámky objevují ve formě bodů i polygonů, proto jsme je stejným postupem jako v předchozím případě převedli na výhradně bodové prvky.

Ověření, zda se vrstvě nenachází 2 body definující 1 objekt, byla provedena u části prvků vizuálně na základě průniku původní polygonové vrstvy s původní bodovou vrstvou. Protože vrstva obsahuje velké množství prvků, navíc často s nevyplněným atributem `name`, a také je obtížné rozhodnout, jestli se v daném místě nenachází několik objektů v těsné blízkosti, nebyla kontrola úplná. Vrstva je přesto validní.

## 3.2 Liniové vrstvy

### 3.2.1 Silnice

Do vrstvy komunikací byly ze schématu `osm` vybrány prvky s atributem `highway` (pozemní komunikace) o hodnotě `motorway` (dálnice), `motorway_link` (nájezdy a sjezdy k dálnicím), `trunk` (rychlostní komunikace), `trunk_link` (nájezdy a sjezdy k rychlostní komunikaci), `primary` (silnice 1. třídy), `primary_link` (přípojně rampy ke komunikaci 1. třídy), `secondary` (silnice 2. třídy), `secondary_link` (přípojně rampy ke komunikaci 2. třídy), `tertiary` (silnice 3. třídy), `tertiary_link` (přípojně rampy ke komunikaci 3. třídy).

Validita vrstvy byla ověřena pomocí funkce `ST_IsValid`<sup>4</sup>, která pro liniové prvky kontroluje, zda jsou jednoduché, čili zda řetězec neprochází 1 bodem 2krát, s výjimkou koncového bodu. Funkce nenalezla žádné invalidní prvky.

```
SELECT *  
FROM silnice  
WHERE NOT ST_IsValid(way)
```

---

<sup>4</sup>[http://postgis.refractor.net/docs/ST\\_IsValid.html](http://postgis.refractor.net/docs/ST_IsValid.html)



## 3.3 Polygonové vrstvy

### 3.3.1 Obce

Vrstva obcí ČR byla převzata ze schématu gis1 a pomocí příkazu převzatého ze stránek předmětu přetransformována ze souřadnicového systému S-JTSK Krovak East North do souřadnicového systému Google Mercator.

## 4 Dotazy

Dotazy jsme se snažili formulovat tak, abychom ukázali určitou praktickou využitelnost databází prostorových dat a zároveň použili různé funkce a postupy.

U každého dotazu uvádíme slovní zadání, SQL příkaz a přibližnou dobu vyhodnocení v programu pgAdmin v1.12.3. Některé prostorové dotazy jsou doplněny vizualizací řešení v programu Quantum GIS 1.6.0

### 4.1 Atributové dotazy

1. Kolik je hradů a zámků, jejichž název končí na 'ice'?

```
SELECT COUNT(name)
FROM hrady_zamky
WHERE name LIKE '%ice'
```

```
count
-----
      47
(1 row)
```

Celková doba provádění dotazu: 11 ms.  
Získán 1 řádek.

2. Kolik obyvatel ČR připadá průměrně na 1 koupaliště?

```
SELECT
(SELECT SUM(obyv02) FROM obce)/COUNT(osm_id)
FROM koupani
```

```
      ?column?
-----
95357.654205607476636
(1 row)
```

Celková doba provádění dotazu: 12 ms.  
Získán 1 řádek.

3. Jaká je hustota hradů na území ČR?

```
SELECT  
COUNT(osm_id)/(SELECT SUM(area)/1e6 FROM obce)  
FROM hrady_zamky
```

?column?

-----

0.00446052726133204921

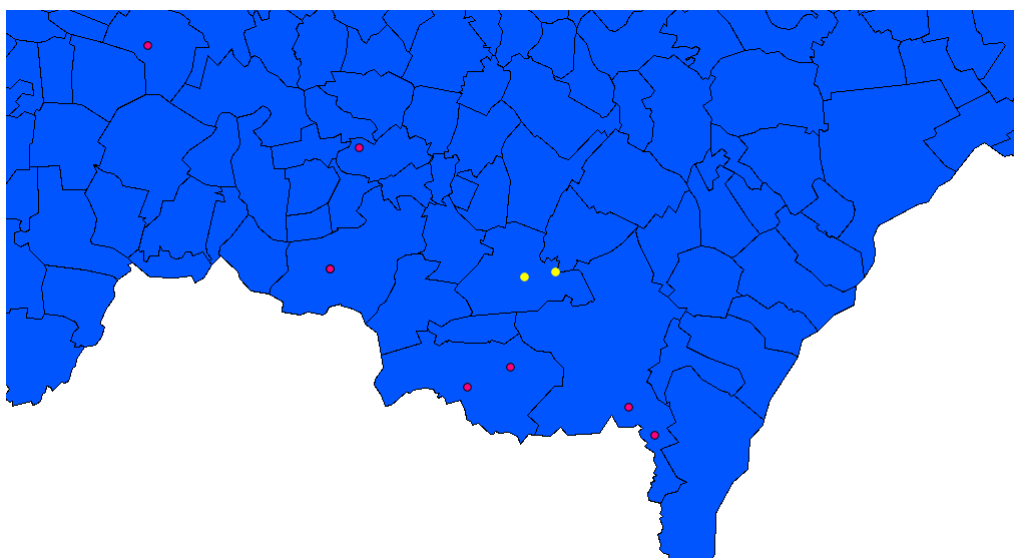
(1 row)

Celková doba provádění dotazu: 23 ms.

Získán 1 řádek.

## 4.2 Prostorové dotazy

1. Které hrady a zámky se nachází na území obce Lednice?



Obr.3: 1. Které hrady a zámky se nachází na území obce Lednice?

```
SELECT h.name  
FROM hrady_zamky AS h  
JOIN obce AS c  
ON ST_Intersects(h.way, c.geom)
```

```
WHERE c.nazev='Lednice'
```

```
name
```

```
-----
```

```
Janův hrad
```

```
Zámek Lednice
```

```
(2 rows)
```

```
Celková doba provádění dotazu: 12 ms.
```

```
Získány 2 řádky.
```

2. Na území které obce je nejvíce hradů a zámků?

```
SELECT c.nazev
FROM hrady_zamky AS h
JOIN obce AS c
ON ST_Intersects(h.way, c.geom)
GROUP BY nazev
ORDER BY count(*) DESC
LIMIT 1
```

```
nazev
```

```
-----
```

```
Praha
```

```
(1 row)
```

```
Celková doba provádění dotazu: 754 ms.
```

```
Získán 1 řádek.
```

3. Které hrady a zámky leží do 200 m od nejbližší silnice na území obce Praha?

```
SELECT DISTINCT(name)
FROM hrady_zamky AS h
JOIN silnice AS s
ON ST_Dwithin(h.way, s.way, 200)
JOIN obce AS o
ON ST_Intersects(h.way, o.geom)
WHERE o.nazev='Praha'
```

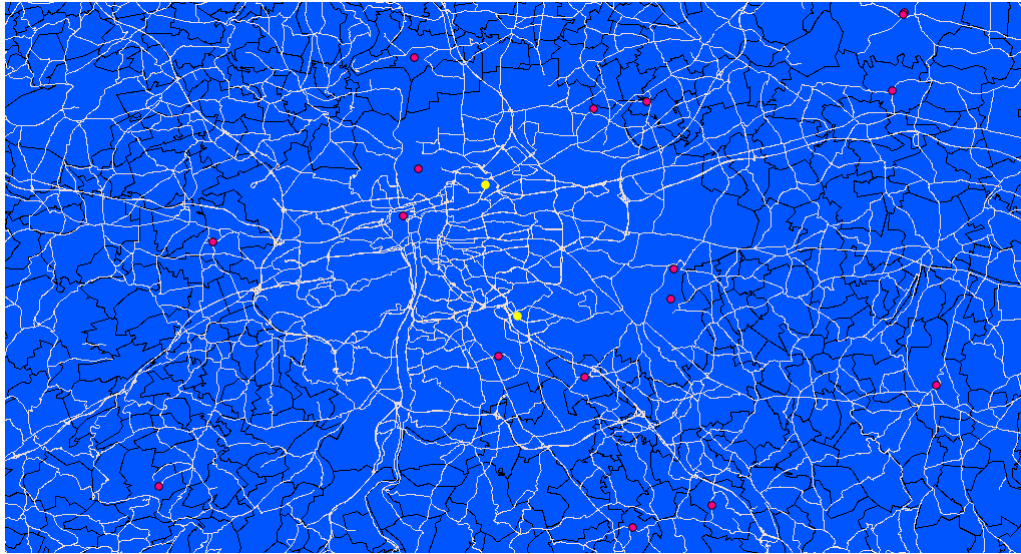
```
name
```

```
-----
```

```
Libeňský zámek
```

```
Chodovská tvrz
```

```
(2 rows)
```



Obr.4: 3. Které hrady a zámky leží do 200 m od nejbližší silnice na území obce Praha?

Celková doba provádění dotazu: 23 ms.  
Získány 2 řádky.

4. Kolik stravovacích zařízení leží do 1 km od hradu Státní hrad a zámek Český Krumlov?



Obr.5: 4. Kolik stravovacích zařízení leží do 1 km od hradu Státní hrad a zámek Český Krumlov?

```

SELECT COUNT(s.name)
FROM hrady_zamky AS h
JOIN stravovani AS s
ON ST_Dwithin(h.way, s.way, 1000)
WHERE h.name='Státní hrad a zámek Český Krumlov'

```

```

count
-----

```

```

      8

```

```

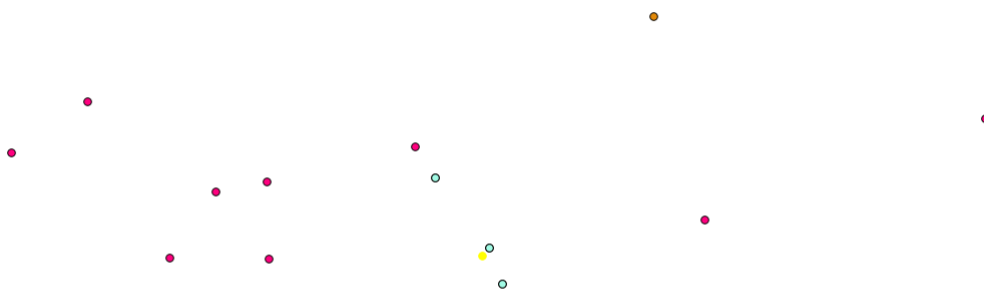
(1 row)

```

Celková doba provádění dotazu: 13 ms.

Získán 1 řádek.

5. Kolik koupališť leží do 10 km od hradu Znojemský hrad?



Obr.6: 5. Kolik koupališť leží do 10 km od hradu Znojemský hrad?

```

SELECT COUNT(k.name)
FROM hrady_zamky AS h
JOIN koupani AS k
ON ST_Dwithin(h.way, k.way, 10000)
WHERE h.name='Znojemský hrad'

```

```

count
-----

```

```

      3

```

```

(1 row)

```

Celková doba provádění dotazu: 24 ms.  
Získán 1 řádek.

6. Kolik stravovacích zařízení leží do 0,5 km od koupaliště Aquacentrum Bospor?



Obr.7: 6. Kolik stravovacích zařízení leží do 0,5 km od koupaliště Aquacentrum Bospor?

```
SELECT COUNT(s.name)
FROM stravovani AS s
JOIN koupani AS k
ON ST_Dwithin(s.way, k.way, 500)
WHERE k.name='Aquacentrum Bospor'
```

```
count
-----
      4
(1 row)
```

Celková doba provádění dotazu: 11 ms.  
Získán 1 řádek.

7. Který hrad leží v největší vzdálenosti od silnice?

```
SELECT h.name
FROM hrady_zamky AS h
```

```
CROSS JOIN silnice AS s
GROUP BY h.name
ORDER BY MIN(ST_Distance(h.way, s.way)) DESC
LIMIT 1
```

```
name
-----
```

```
Lány
(1 row)
```

Celková doba provádění dotazu: 43085 ms.  
Získán 1 řádek.

8. Který hrad leží v nejmenší vzdálenosti od silnice?

```
SELECT h.name
FROM hrady_zamky AS h
CROSS JOIN silnice AS s
ORDER BY ST_Distance(h.way, s.way) ASC
LIMIT 1
```

```
name
-----
```

```
Moravské Budějovice
(1 row)
```

Celková doba provádění dotazu: 42753 ms.  
Získán 1 řádek.

9. Které stravovací zařízení leží nejbliže hradu Svojanov?

```
SELECT s.name
FROM stravovani AS s
CROSS JOIN hrady_zamky AS h
WHERE h.name='Svojanov'
ORDER BY ST_Distance(s.way, h.way)
LIMIT 1
```

```
name
-----
```

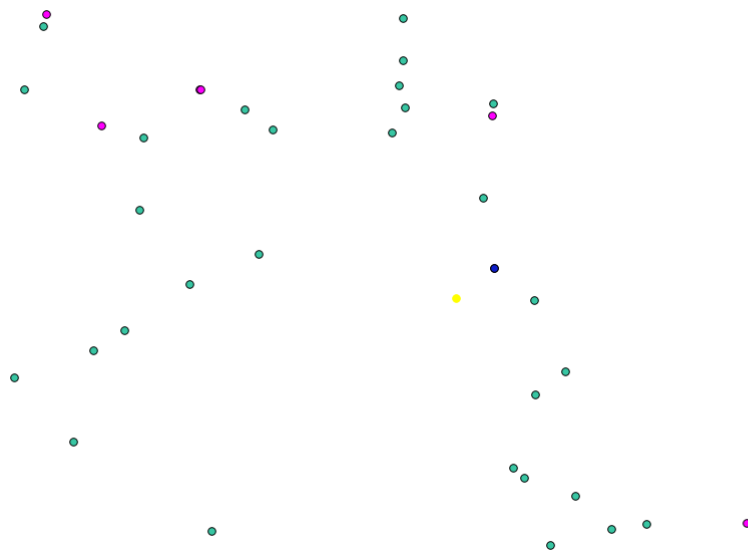
```
Na Hrdé Vsi
(1 row)
```

Celková doba provádění dotazu: 22 ms.  
Získán 1 řádek.



Obr.8: 9. Které stravovací zařízení leží nejbližze hradu Svojanov?

10. Které koupaliště leží nejbližze hradu Zámek Opočno?



Obr.9: 10. Které koupaliště leží nejbližze hradu Zámek Opočno?

```
SELECT k.name
FROM koupani AS k
CROSS JOIN hrady_zamky AS h
WHERE h.name='Zámek Opočno'
ORDER BY ST_Distance(k.way, h.way)
```



LIMIT 1

```
      name
-----
```

Dobrušský bazén

(1 row)

Celková doba provádění dotazu: 12 ms.

Získán 1 řádek.

11. Na území které obce (nazev, gid) je největší hustota hradů (nejvíce hradů na km<sup>2</sup>)?

```
SELECT nazev, gid
FROM hrady_zamky
JOIN obce
ON ST_Intersects(geom, way)
GROUP BY area, nazev, gid
ORDER BY COUNT(osm_id)/area DESC
LIMIT 1
```

```
      nazev      | gid
-----+-----
```

Josefův Důl | 758

(1 row)

Celková doba provádění dotazu: 73 ms.

Získán 1 řádek.

12. V které obci je nejvíce stravovacích zařízení na obyvatele?

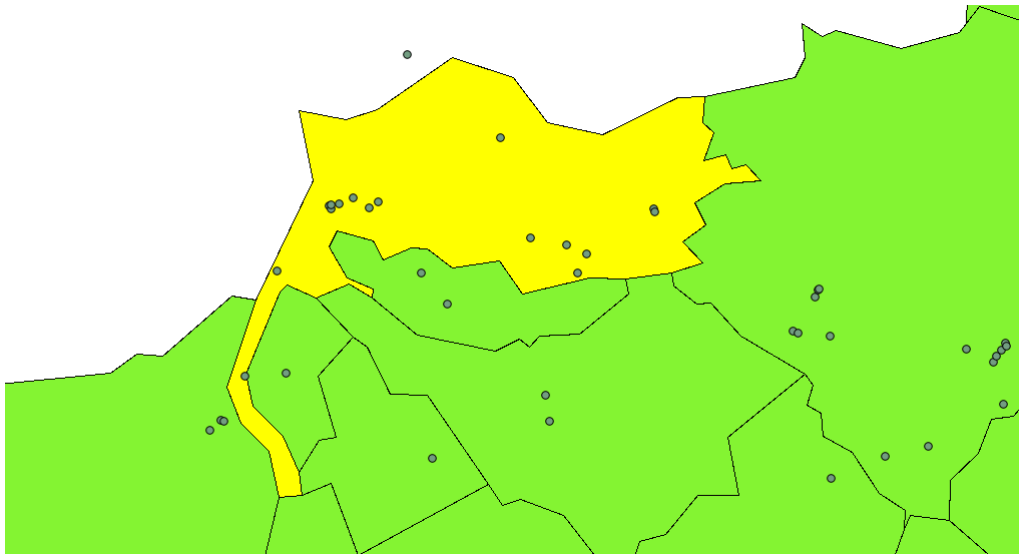
```
SELECT nazev
FROM stravovani
JOIN obce
ON ST_Intersects(geom, way)
GROUP BY obyv02, nazev
ORDER BY COUNT(osm_id)/obyv02 DESC
LIMIT 1
```

```
      nazev
-----
```

Hřensko

(1 row)

Celková doba provádění dotazu: 285 ms.



Obr.10: 12. V které obci je nejvíce stravovacích zařízení na obyvatele?

Získán 1 řádek.

13. Kolik stravovacích zařízení leží do 100 m od dálnice?

```
SELECT COUNT(h.name)
FROM stravovani AS h
JOIN silnice AS s
ON ST_Dwithin(h.way, s.way, 100)
WHERE typ='motorway'
```

```
count
-----
      14
(1 row)
```

Celková doba provádění dotazu: 52 ms.

Získán 1 řádek.

14. Která pizzerie (střávková zařízení, jehož název obsahuje řetězec izz) leží nejbližšího hradu?

```
SELECT s.name, h.name
FROM stravovani AS s
CROSS JOIN hrady_zamky AS h
WHERE s.name LIKE '%izz%'
```

```
ORDER BY ST_Distance(s.way, h.way)
LIMIT 1
```

```
      name      |      name
-----+-----
 Pizzerie Renardo | Vrchlabský zámek
(1 row)
Celková doba provádění dotazu: 71 ms.
Získán 1 řádek.
```

15. V kolika obcích nejsou žádná stravovací zařízení?

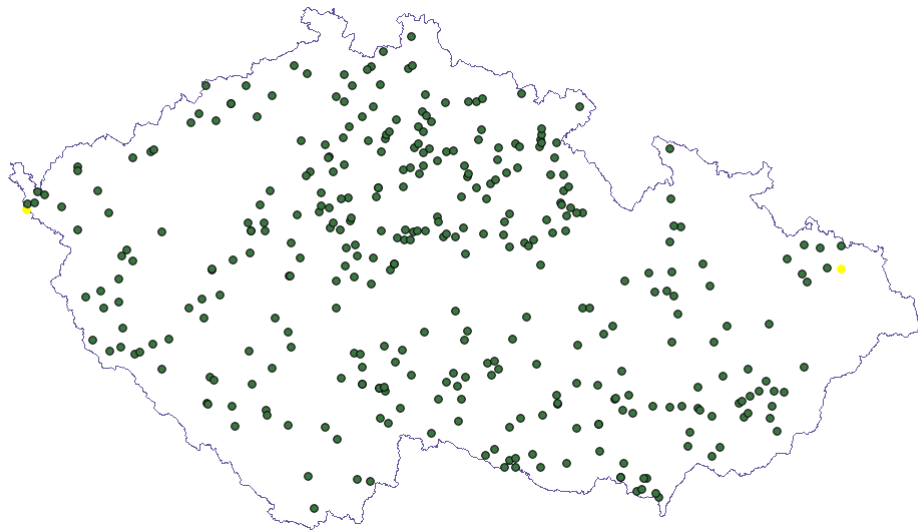
```
SELECT COUNT(DISTINCT(gid))
FROM obce w
WHERE gid NOT IN
(SELECT DISTINCT(gid)
FROM stravovani
JOIN obce
ON ST_Intersects(way, geom))
```

```
count
-----
 5270
(1 row)
Celková doba provádění dotazu: 276 ms.
Získán 1 řádek.
```

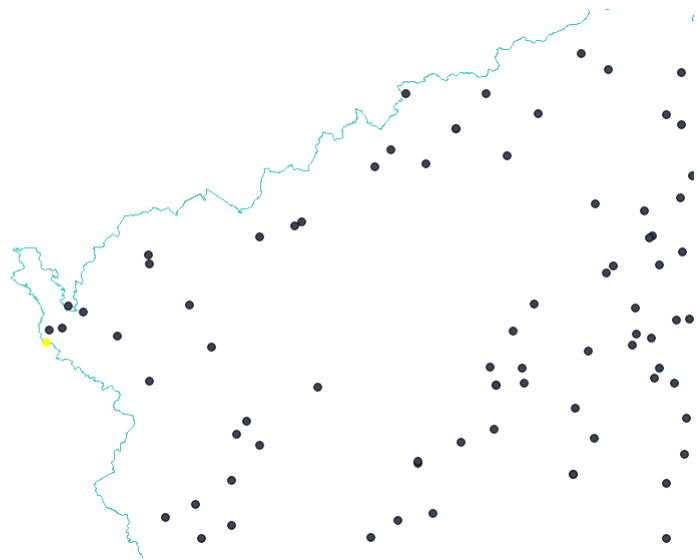
16. Které 2 hrady jsou od sebe nejdál?

```
SELECT h.name, z.name
FROM hrady_zamky AS h
CROSS JOIN hrady_zamky AS z
ORDER BY ST_Distance(h.way, z.way) DESC
LIMIT 1
```

```
      name      |      name
-----+-----
Rothschildův empírový zámek | Burg Hohenberg
(1 row)
Celková doba provádění dotazu: 173 ms.
Získán 1 řádek.
```



Obr.11: 16. Které 2 hrady jsou od sebe nejvzdálenější?



Obr.12: 17. Který hrad leží nejbližě hranicím ČR?

17. Který hrad leží nejbližě hranicím ČR?

```
DROP TABLE hranice;
```

```
CREATE TABLE hranice
AS SELECT ST_Boundary(ST_Union(geom)) AS way
FROM obce;
```

```

SELECT h.name
FROM hranice AS b
JOIN hrady_zamky AS h
ON b.way && h.way
ORDER BY ST_Distance(b.way, h.way)
LIMIT 1

```

```

          name
-----
    Burg Hohenberg
(1 row)
Celková doba provádění dotazu: 73 ms.
Získán 1 řádek.

```

## 5 Použitý software

- nástroje pro administraci a správu PostgreSQL **pgAdmin v1.10.5**
- svobodný a multiplatformní geografický informační systém **Quantum GIS 1.4.0**
- sázecí program **L<sup>A</sup>T<sub>E</sub>X** a kompilátor **MiKTeX 2.8**

## 6 Závěr

Výsledkem našeho semestrálního projektu v rámci předmětu Úvod do zpracování prostorových dat je 5 validních tematických vrstev (*hrady\_zamky*, *koupani*, *stravovani*, *silnice*, *obce*) ve schématu *e11* v databázi *pgis\_uzpd* a sada atributových a prostorových dotazů, které byly nad těmito vrstvami provedeny. Jak moc jejich výsledky odpovídají skutečnosti, závisí především na kvalitě vstupních dat z projektu OpenStreetMap, jež se pro různé části území České republiky liší.

## 7 Seznam příloh

- SQL dávkový soubor
- Prezentace

## Reference

- [1] 153UZPD Úvod do zpracování prostorových dat  
<http://geo101.fsv.cvut.cz/gwiki/153UZPD>

- [2] PostGIS 1.5.2 Manual  
<http://postgis.refrations.net/docs/>
- [3] projekt OpenStreetMap  
<http://www.openstreetmap.org/>
- [4] software pgAdmin  
<http://www.pgadmin.org/>
- [5] software Quantum GIS  
<http://www.qgis.org/>