

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Katedra geomatiky



Úvod do zpracování prostorových dat

Semestrální projekt

## Obsah

1.	Zadání .....	1
2.	Úvod .....	1
2.1	Výběr tématu.....	1
2.2	Použitý software.....	1
3	Zdroje dat .....	1
3.1	Český statistický úřad .....	1
4	Struktura dat .....	2
5	Dotazy a bodování.....	4
6	Závěr .....	6

## 1. Zadání

Navrhňte a vytvořte tematické vrstvy (např. vodní toky, vodní plochy, lesy, silnice, železnice a pod.) na základě dat OpenStreetMap (viz schéma osm) 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\_uzpr.

## 2. Úvod

### 2.1 Výběr tématu

Cílem semestrálního projektu bylo vytvoření databáze a SQL dotazů pro SQL tutor.

### 2.2 Použitý software

QGIS – Zdarma a otevřený zdrojový geografický informační systém verze 3.10.2

## 3 Zdroje dat

### 3.1 Český statistický úřad

Český statistický úřad je ústřední orgán státní správy České republiky. Byl zřízen dne 8. ledna 1969 zákonem č. 2/1969 Sb., o zřízení ministerstev a jiných ústředních orgánů státní správy. Působnost ČSÚ je vymezena zákonem č. 89/1995 Sb., o státní statistické službě. Český statistický úřad zabezpečuje získávání a zpracování údajů pro statistické účely a poskytuje statistické informace státním orgánům, orgánům územní samosprávy, veřejnosti a do zahraničí. Zajišťuje vzájemnou srovnatelnost statistických informací ve vnitrostátním i mezinárodním měřítku. K tomu stanoví metodiku statistických zjišťování včetně programu statistických zjišťování. Uvedený program přesně vymezuje působnost ČSÚ a dalších rezortů, v jejichž náplni je rovněž statistické zjišťování. Do státního statistického systému lze zařadit nejen výsledky statistických zjišťování, nýbrž i administrativní údaje, které jsou významným zdrojem statistických informací. Nedílnou součástí statistického systému je statistický metainformační systém, který díky unikátnímu popisu ukazatelů přesně definuje jednotlivé ukazatele používané a využívané ve statistice. Metainformace jsou standardním prvkem všech údajů zveřejněných ve veřejné databázi ČSÚ. ČSÚ zajišťuje zpracování i zveřejňování údajů, sestavuje souhrnné statistické charakteristiky vývoje národního hospodářství, zpracovává analýzy, zpracovává projekce demografického vývoje, provádí konjunkturální průzkumy. Sleduje dodržování povinností v oblasti státní statistické služby, vytváří a spravuje statistické klasifikace, statistické číselníky a statistické registry a poskytuje z nich informace, poskytuje statistické informace do zahraničí, spolupracuje s mezinárodními organizacemi, podílí se na tvorbě statistiky Evropských společenství. ČSÚ vykonává státní správu na úseku zpracování výsledků voleb do zastupitelstev územních samosprávných celků, Parlamentu České republiky a do Evropského parlamentu konaných na území

České republiky a zpracování výsledků celostátního referenda v rozsahu stanoveném zvláštními právními předpisy. Výsledky voleb jsou on-line zveřejňovány na [www.volby.cz](http://www.volby.cz). Jednou za 10 let provádí Sčítání lidu, domů a bytů, jehož výsledky jsou pak nejbohatším zdrojem informací o obyvatelstvu. Poslední sčítání bylo uskutečněno v roce 2011. Český statistický úřad dále shromažďuje zahraniční statistické informace, zejména pro účely porovnání úrovně sociálního, ekonomického, demografického a ekologického vývoje státu se zahraničím. Získané informace i výsledky porovnání poskytuje státním orgánům i veřejnosti. Veškerá data a informace jsou zdarma dostupné na [www.czso.cz](http://www.czso.cz). Vydává Statistickou ročenku České republiky, další odborné publikace a časopisy a zajišťuje vývoj nových statistických metod. Nejnovější údaje o ekonomickém i sociálním vývoji ČR se zveřejňují formou tzv. Rychlých informací. Dále pak na svých stránkách zveřejňuje analýzy, on-line publikace, datové sady, časové řady, tiskové zprávy a jiné informace. [1]

## 4 Struktura dat

Pro projekt byla vybrána data z voleb do poslanecké sněmovny 2017. Tabulky jsou vytvořeny SQL dávkou *davka.sql*.

kraje	
náz. sloupce	typ proměnné
id	INT PRIMARY KEY
nazev	VARCHAR(50)
rozloha	INT
obyvatele	INT

Tabulka *kraje* obsahuje 14 záznamů shrnujících obecné informace o krajích.

strany	
náz. sloupce	typ proměnné
id	INT PRIMARY KEY
nazev	VARCHAR(200)
leader	VARCHAR(40)
zalozeni	INT

Tabulka *strany* obsahuje 31 záznamů shrnujících obecné informace o volebních stranách.

vysledky	
náz. sloupce	typ proměnné
id	INT PRIMARY KEY
id_kraj	INT
id_strana	INT
platne_hlasy	INT
počet_mandaty	INT

Tabulka *vysledky* obsahuje 434 záznamů. Výsledky jsou uváděny pro konkrétní stranu v konkrétním kraji.

Sloupce určené pro propojení tabulek:

-*id\_kraj* – odpovídá sloupci *id* v tabulce *kraje*

-*id\_strana* – odpovídá sloupci *id* v tabulce *strany*

zakl_info	
náz. sloupce	typ proměnné
id_kraj	INT
pocet_okrsku	INT
počet_volici	INT
hlasujici_volici	INT
platne_hlasy	INT
kandidati_20_39	INT
kandidati_40_59	INT
kandidati_nad_60	INT
kandidati_prum_vek	FLOAT
zvoleni_20_39	INT
zvoleni_40_59	INT
zvoleni_nad_60	INT
zvoleni_prum_vek	FLOAT

Tabulka *zakl\_info* obsahuje 14 záznamů podle jednotlivých krajů. Informace zahrnují např. Počty voličů, platných hlasů, kandidátů a zvolených kandidátů různých věkových skupin.

Sloupce určené pro propojení tabulek:

-*id\_kraj* – odpovídá sloupci *id* v tabulce *kraje*

## 5 Dotazy a bodování

Jednotlivé dotazy pro SQL tutor byly sepsány do textového souboru dotazy.txt ve formátu:

```
SELECT insert_problem ("název datasetu", "číslo otázky", "bodové ohodnocení", 'select');
```

```
SELECT insert_question("název datasetu", "číslo otázky", "pořadí otázek", 'cs',  
"Text otázky");
```

```
SELECT insert_answer ("název datasetu", "číslo otázky", "pořadí odpovědi",  
"SQL řešení");
```

Hodnocení dotazů bylo přiřazováno na základě použití speciálních funkcí, propojení tabulek nebo komplikovanější úvahy.

Příklad č. 1 (Řešení lze nalézt v rámci jedné tabulky za použití klasických příkazů. (1 bod)):

```
SELECT insert_problem ('volby', 602, 1, 'select');
```

```
SELECT insert_question('volby', 602, 1, 'cs',
```

```
'Vypište rok založení a název strany, jejíž leader je "Jan Hamáček"?');
```

```
SELECT insert_answer ('volby', 602, 1,
```

```
'SELECT zalozeni, nazev
```

```
FROM strany
```

```
WHERE leader='Jan Hamáček');
```

Příklad č. 2 (Zde bylo třeba propojení dvou tabulek, agregační funkce *sum* a logického operátoru *OR*. (3 body)):

```
SELECT insert_problem ('volby', 605, 3, 'select');
```

```
SELECT insert_question('volby', 605, 1, 'cs',
```

```
'Kolik lidí mělo možnost volit v krajích "Liberecký kraj" a "Královéhradecký kraj"?');
```

```
SELECT insert_answer ('volby', 605, 1,
```

```
'SELECT sum(pocet_volici)
```

```
FROM zakl_info
```

```
JOIN kraje ON kraje.id = zakl_info.id_kraj
```

```
WHERE kraje.nazev = 'Liberecký kraj' OR kraje.nazev = 'Královéhradecký kraj');
```

Příklad č. 3 (Kromě propojení tabulek a agregační funkce jsou zde využity příkazy GROUP BY, ORDER BY a LIMIT (4 body)):

```
SELECT insert_problem ('volby', 615, 4, 'select');
SELECT insert_question('volby', 615, 1, 'cs',
'Které 3 strany získaly nejvíce mandátů? Vypište názvy stran a počty mandátů. ');
SELECT insert_answer ('volby', 615, 1,
'SELECT strany.nazev, sum(pocet_mandaty)
FROM strany
JOIN vysledky ON vysledky.id_strana = strany.id
GROUP BY strany.nazev
ORDER BY sum(pocet_mandaty) DESC
LIMIT 3;');
```

Příklad č. 4 (Zde je navíc nutné propojit 3 tabulky a subselect (6 bodů)):

```
SELECT insert_problem ('volby', 619, 6, 'select');
SELECT insert_question('volby', 619, 1, 'cs',
'Která strana získala nejvíce mandátů v největším kraji? Vypište stranu, kraj a počet mandátů. ');
SELECT insert_answer ('volby', 619, 1,
'SELECT strany.nazev, kraje.nazev, pocet_mandaty
FROM vysledky
JOIN strany ON vysledky.id_strana = strany.id
JOIN kraje ON vysledky.id_kraj = kraje.id
WHERE id_kraj =
(SELECT id
FROM kraje
ORDER BY rozloha DESC
LIMIT 1)
ORDER BY pocet_mandaty DESC
LIMIT 1;');
```

## 6 Závěr

Cíl projektu byl splněn a byly vytvořené všechny podklady pro výukovou sadu *volby*.

Návrh na rozšíření:

Stávající tabulky by bylo možné rozšířit o data z voleb do poslanecké sněmovny z minulých let, pokud by byla ve stejném formátu. Pravděpodobně by bylo třeba pozměnit strukturu dat a přidat další tabulku s označením jednotlivých voleb.

[1] Český statistický úřad – Wikipedie. [online]. Dostupné z:  
[https://cs.wikipedia.org/wiki/%C4%8Cesk%C3%BD\\_statistick%C3%BD\\_%C3%BA%C5%99ad](https://cs.wikipedia.org/wiki/%C4%8Cesk%C3%BD_statistick%C3%BD_%C3%BA%C5%99ad)



