

VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA

Hornicko-geologická fakulta

Institut geoinformatiky

MONITORING CEN NEMOVITOSTÍ Z INZERCE REALITNÍHO PORTÁLU

Diplomová práce

Autor: Bc. Martin Matuszczyk

Vedoucí diplomové práce: Ing. Igor Ivan, Ph.D.

Ostrava 2014

VŠB - Technická univerzita Ostrava
Hornicko-geologická fakulta
Institut geoinformatiky

Zadání diplomové práce

Student: **Bc. Martin Matuszczyk**
Studijní program: N3654 Geodézie, kartografie a geoinformatika
Studijní obor: 3608T002 Geoinformatika
Téma: **Monitoring cen nemovitostí z inzerce realitního portálu**
Monitoring of Price of Real Estates from Real Estate Portal

Zásady pro vypracování:

1. Rešerše literatury popisující podobná řešení
2. Výběr vhodného realitního serveru
3. Návrh a realizace databáze
4. Inovace postupu pro stahování a import dat do databáze s využitím RSS kanálu
5. Generalizace a vizualizace stažených dat; příprava pro budoucí aktualizace
6. Publikace výsledků pro vybrané území na webu

Rozsah grafických prací:
dle potřeby

Rozsah původní zprávy:
30 - 50 stran textu

Seznam doporučené odborné literatury:

- Longley P. A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. (2010): Geographic Information Systems and Science, Wiley, 560 s.
Voženilek, V., Kaňok, J. a kol. (2011): Metody tematické kartografie. Univerzita Palackého v Olomouci, Olomouc, 216 s.
O'Sullivan, D., Unwin, D. (2010): Geographic Information Analysis. Wiley, 432 s.


Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Igor Ivan, Ph.D.**

Datum zadání: 31.10.2013
Datum odevzdání: 30.04.2014



prof. Ing. Zdeněk Diviš, CSc.
vedoucí institutu



prof. Ing. Vladimír Slivka, CSc., dr.h.c.
děkan fakulty

Prohlášení

- Celou diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.
- Byl jsem byl seznámen s tím, že na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo.
- Beru na vědomí, že Vysoká škola báňská – Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užít (§ 35 odst. 3).
- Souhlasím s tím, že jeden výtisk diplomové práce bude uložen v Ústřední knihovně VŠB-TUO k prezenčnímu nahlédnutí a jeden výtisk bude uložen u vedoucího diplomové práce. Souhlasím s tím, že údaje o diplomové práci, obsažené v Záznamu o závěrečné práci, umístěném v příloze mé diplomové práce, budou zveřejněny v informačním systému VŠB-TUO.
- Souhlasím s tím, že diplomová práce je licencována pod Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licencí. Pro zobrazení kopie této licence, je možno navštívit <http://creativecommons.org/licenses/by-nc-sa/3.0/>
- Bylo sjednáno, že s VŠB-TUO, v případě zájmu o komerční využití z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona.
- Bylo sjednáno, že užít své dílo – diplomovou práci nebo poskytnout licenci k jejímu komerčnímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Ostravě dne

Martin Matuszczyk

Anotace

Diplomová práce se zabývá monitoringem cen nemovitostí prostřednictvím vytvoření cenové mapy prodeje a pronájmu z inzerce vybraného realitního portálu na území České republiky. Inzertní údaje jsou stahovány prostřednictvím RSS kanálu daného portálu, filtrovány a zpracovány tak, aby je bylo možné zobrazit ve webové mapové aplikaci, v této práci vyvinuté. V práci jsou rovněž zmíněny nevýhody daného zdroje dat, ale i srovnání se stávajícími aplikacemi.

Klíčová slova:

Cenová mapa, RSS, OpenLayers, GeoServer, nemovitosti

Summary

This thesis deals with monitoring of real estate prices by creating a price map of sales and leases from advertising of selected real estate portal in the Czech Republic. Advertisements are downloaded through the RSS feed, filtered and processed, so they can be shown in a web map application developed in this thesis. The thesis also mentions disadvantages of chosen data source and compares final application with existing applications in Czech Republic.

Keywords:

Price map, RSS, OpenLayers, GeoServer, real estates

Rád bych tímto poděkoval Ing. Igoru Ivanovi, Ph.D. za jeho konstruktivní a pečlivé připomínky, díky kterým má práce dospěla do kýženého cíle. Dále pak Ing. Janu Růžičkovi, Ph.D. a Ing. Davidu Vojtkovi, Ph.D. za zprostředkování školního serveru pro finální aplikaci a Institutu Geoinformatiky za poskytnutí potřebných znalostí a licencí software, bez kterých by nebylo možné tuto práci vypracovat.

Obsah

1. Úvod.....	1
2. Cíle.....	2
3. Rešerše podobných řešení.....	3
3.1. Cenové mapy na internetu.....	4
4. Výběr vhodného realitního serveru.....	10
4.1. Popis realitního serveru Sreality.....	10
4.2. Popis RSS kanálu sreality.cz.....	11
4.3. Žádost o spolupráci s portálem seznam.cz.....	13
5. Návrh a realizace databáze.....	14
5.1. Tabulky pro uložení stažených dat.....	15
5.2. Podpůrné tabulky pro prostorovou reprezentaci.....	16
5.3. Struktura databáze.....	18
6. Inovace postupu pro stahování a import dat do databáze.....	19
6.1. Kód skriptu.....	20
6.2. Automatizace stahování.....	22
7. Filtrace, agregace a export dat.....	23
7.1. Filtrace dat.....	23
7.2. Agregace a výběr dat pro export.....	25
7.3. Export dat.....	27
8. Výběr vhodné metody zobrazení, automatizace zpracování.....	29
8.1. Zobrazení pomocí kartogramů a ohodnocených linií.....	29
8.2. Zobrazení pomocí rastrů.....	33
8.3. Automatizace propojení dat.....	35
8.4. Nedostatky plynoucí ze zdroje dat.....	36
9. Tvorba webové mapové aplikace.....	37
9.1. Publikace vrstev prostřednictvím aplikace Geoserver.....	37
9.2. Vývoj aplikace s užitím knihoven OpenLayers.....	40
9.3. Porovnání aplikace s konkurencí.....	43
10. Závěr.....	45
Seznam zdrojů.....	46
Seznam obrázků.....	47
Seznam tabulek.....	47
Seznam příloh na CD.....	47

1. Úvod

Situace na českém trhu nemovitostí disponuje jedním podstatným nedostatkem. Tím je zkrácení cen nemovitostí, které ovlivňuje trh jak ze strany realitních kanceláří a jejich klientů, tak ze strany bank a developerů. Tento nedostatek zásadně ovlivňuje vztahy mezi klienty, usilujícími o koupi či prodej nemovitosti, bankami, které jim pro koupi nemovitosti poskytují hypotéky a developery, nabízejícími nové nemovitosti, či realitními kancelářemi usilujícími o prodej nemovitostí stávajících.

Základem tohoto nedostatku je subjektivní názor zúčastněných na cenu nemovitostí. Tento názor se totiž na cenu jedné a té samé nemovitosti ze strany klientů, realitních kanceláří, developerů a bank často velmi liší. Realitní kanceláře a developery na trhu prezentují určité nabídkové ceny, tedy ceny za které chtějí dané nemovitosti prodat. Banky pro poskytnutí hypotéky vytváří pojem obvyklé ceny závislé na subjektivních kritériích dané banky a odhadu vybraného znalce. Klienti pak ztrácí přehled o cenách reálných a přistupují na ceny jimi stanovené.

Rozdílnost názoru na cenu je navíc dána relativní nevyzrálostí moderního trhu nemovitostí. Ten vzniknul teprve v polovině devadesátých let, kdy byla ukončena centrálně plánovaná výstavba panelových sídlišť. Poté se trh začal ubírat více komerční cestou, bez toho aby si prošel tradičním vývojem. Situace ohledně pronájmu bytů je pak výrazně ovlivněna deregulací nájemného.

Mládí trhu nových nemovitostí a rozdíly v názoru na jejich cenu měly za následek nadměrně pozitivní očekávání při růstu cen v letech 2005 – 2008, ale i skepsi k cenám nemovitostí v letech 2010 – 2011, tedy v době dopadu finanční krize na realitní trh. První případ byl charakteristický neúměrným zvyšováním cen a poptávkou, zatímco druhý čekáním na pokles cen a tedy snížením poptávkou.

Tyto události dospěly do takového rozsahu, že zájemci o koupi, prodej či pronájem nemovitosti, neorientující se v současných trendech trhu s nemovitostmi, neví, jaké částky požadovat. Z této situace pak vyplynula snaha o pomoc těmto zájemcům formou určitého monitoringu trhu nemovitostí a vyvinutí aplikace cenové mapy umožňující jednodušší orientaci v současných cenách na trhu nemovitostí, ať už se jedná o prodej či pronájem, prostřednictvím této diplomové práce.

Překážkou této snahy je však skutečnost, že reálně uskutečněné ceny prodeje a pronájmů nejsou veřejně přístupné, a proto je potřeba využít určitou substituci požadovaného datového zdroje, tedy inzertní nabídky nemovitostí.

2. Cíle

Hlavním cílem této diplomové práce, jak již bylo nastíněno v úvodu, je vytvoření webové aplikace cenové mapy. Ta by měla poskytnout koncovým uživatelům informaci o cenách prodeje, či pronájmu nemovitostí v zájmové lokalitě. Jako každá mapa, měla by i tato aplikace splňovat všechny kartografické požadavky a obsahovat základní kompoziční prvky. Bylo by taktéž vhodné nalézt inspiraci ve stávajících projektech cenových map a webových aplikací.

Aby bylo možné vyvinout takovou aplikaci, je nutné nalézt vhodný zdroj dat poskytující informace o cenách nemovitostí. Takovým zdrojem by měl být určitý realitní portál umožňující sledování inzertních nabídek prostřednictvím RSS kanálu. Nabídky z daného kanálu by pak měly být automatizovaně stahovány a ukládány do prostředí databáze. Z toho tedy vyplývá úkol vyvinutí metody zpracování daného RSS kanálu a vytvoření požadované databáze.

Údaje z inzerce ukládané do prostředí databáze budou v rámci této diplomové práce sloužit k vytvoření mapových vrstev prezentovaných ve finální aplikaci. Zároveň však mohou být zdrojem dat k jiným projektům a analýzám realizovaným Institutem Geoinformatiky. Proto by měly být v databázi nadále archivovány a aktualizovány.

Během zpracování bude nutné provést výběr vhodné kartografické metody zobrazení nasbíraných dat. Takový výběr by měl být inspirován právě rešerší stávajících projektů a důvody výběru dané metody by měly být stručně popsány.

Po dokončení technické části činností by měla finální aplikace být srovnána se svou konkurencí a mělo by dojít k hodnocení užitečnosti dané aplikace pro koncové uživatele.

3. Rešerše podobných řešení

Důležitým úkolem při zpracování diplomové práce je samozřejmě rešerše literatury a stávajících aplikací zabývajících se podobnou problematikou. Tištěný zdroj, zabývající se přímo hromadným stahování inzertních nabídek, jejich následným zpracováním a vizualizací v prostoru prostřednictvím webové mapové aplikace nebyl nalezen. Byly tedy vyhledány zdroje zabývající se jednotlivými fragmenty této problematiky.

Obecný pohled na užití RSS kanálu jako nového datového zdroje prostorových dat lze nalézt v publikaci *Nové internetové datové zdroje*. [2] Publikace široce pojednává o možnostech sběru, ukládání, agregace a analýz informací publikovaných na internetu. Mimo jiné se autor zmiňuje o možnosti analýz RSS zpráv. Generování vlastní báze dat z dat shromážděných automatizovanou činností na internetu je zde popisováno jako aktivní vytváření dat z informací a služeb na internetu. Proces automatizace je pak řešen vytvořením vlastního robota, který sbírá informace ze specifikovaných internetových zdrojů. Zdůrazňuje taktéž problém řešení aspektů legálnosti a etiky. Při této činnosti nesmí docházet k porušování práv subjektu, jehož data či stránky jsou využívány. Doporučuje tedy předchozí domluvu s poskytovatelem primárních dat a služeb. Zároveň lze dle autora předpokládat, že pokud jsou data využita pro jiný účel, než je primární důvod činnosti poskytovatele dat, neměl by být s takovým využitím problém.

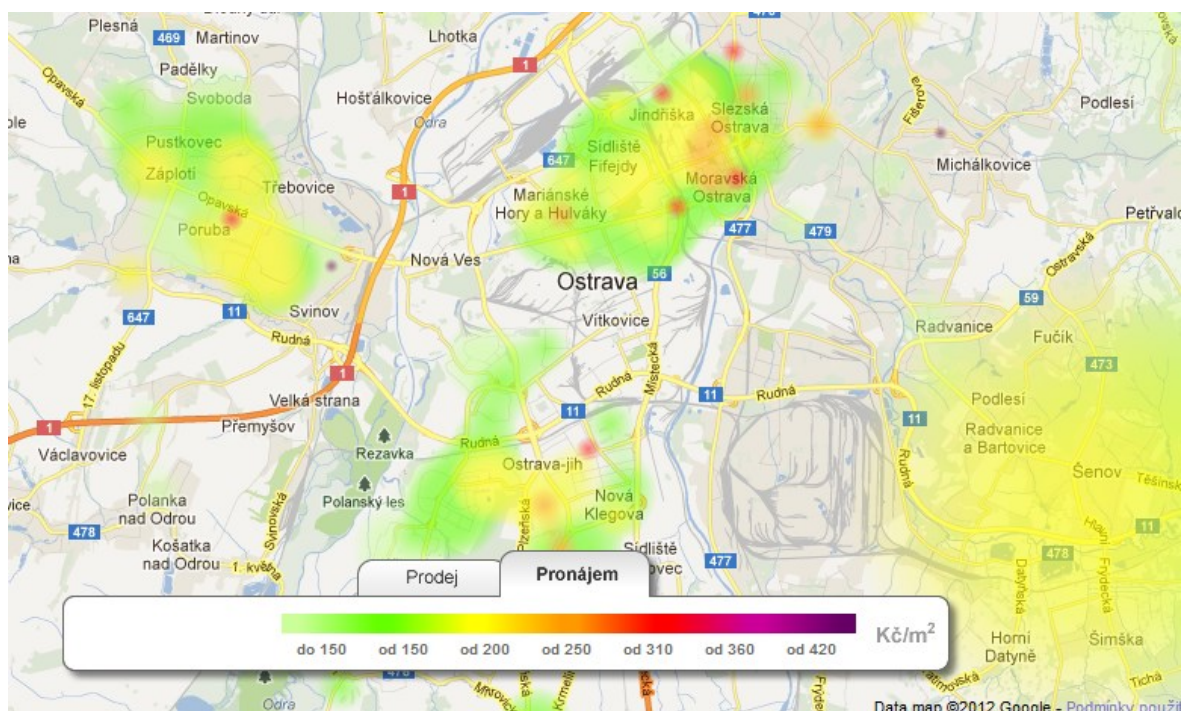
Taktéž je vhodné nahlédnout na povahu samotných dat v diplomové práci užitých. Zdrojem dat jsou v tomto případě inzertní nabídky mnoha realitních kanceláří získané prostřednictvím realitního portálu. Bylo by nerozumné hodnoty cen daných nemovitostí považovat za skutečné transakční ceny. Ceny v inzerci mohou být samozřejmě nadhodnoceny a jejich hodnota častokrát poklesne, než se daná nemovitost prodá. Pro potřeby určení skutečné ceny se používají koeficienty zohledňující zdroj ceny. Tyto si musí každý odhadce stanovit sám pro danou lokalitu, typ nemovitosti, realitní kancelář a druh inzerce. Je tedy zřejmé, že automatizovaně tento problém nelze efektivně řešit. Ojedinele a ve výjimečných případech je možné užit hodnotu 0,85, která by znamenala, že nemovitosti od započetí své inzerce, po samotný prodej poklesnou na své ceně o 15%. [3] V této práci se však zabýváme monitoringem cen z inzerce, nikoliv cen uskutečněných, takže k podobným úpravám pomocí zohledňujících koeficientů nedojde. Problematika je však zmíněna, aby nedocházelo k nedorozuměním, jaké údaje budou ve finální aplikaci prezentovány.

Jelikož je práce zaměřena na vývoj webové aplikace, tak se studium cenových map orientuje na aktivní webové mapové aplikace v další kapitole. Pro potřeby vývoje takové aplikace pak byly prostudovány knihy zabývající se

problematikou OpenLayers a Geoserveru, která je popsána v příslušných kapitolách. [4] [5]

3.1. Cenové mapy na internetu

Vzhledem k povaze diplomové práce je průzkum stávajících cenových map v rámci České republiky absolutní nutností. Jedním z důvodů je samozřejmě inspirace z již fungujících projektů. Ať už se jedná o samotný kartografický vzhled daných aplikací, jejich funkčnost, či zpracování vstupních dat, pokud je tak uvedeno. Dalším důvodem pak je identifikace slabých stránek stávajících cenových map a z toho plynoucí pokus o jejich eliminaci při vytváření vlastního projektu. Celkem byly nalezeny tři české aplikace a jedna zahraniční: cenová mapa ze serveru bezreality.cz, cenová mapa Asociace realitních kancelářů, cenová mapa Státního fondu rozvoje bydlení a Národní stránka cen domovů ze serveru Trulia se zájmovým územím Spojených států.



Obrázek 1: Cenová mapa bezreality.cz

Jako první je zde uvedena cenová mapa ze serveru bezreality.cz. [6] Její vzhled a obsah je zobrazen na obrázku 1 s přiblížením na zájmovou oblast území Ostravy a jejího blízkého okolí. Jedná se o rastrovou mapu vzešlou z interpolace vstupních hodnot přesněji nedefinovanou metodou. Navíc zde není uveden jakýkoliv popis zdrojových dat, časový rámec a postup zpracování. U této aplikace není uvedeno, zda zpracovává pouze byty, nebo i domy. Pokud by však v hodnotách na mapě prodeje byly zahrnuty ceny za domy i byty dohromady,

mohlo by docházet k výraznému zkreslení cen. Každopádně přepínač pro zobrazení cen bytů, či domu chybí. Za chvályhodnou může být uvedena legenda popisující hladiny cen normovaných na metr čtvereční. Legenda skládající se z více barevných tónů a plynulých přechodů mezi nimi zaručuje jednoduché a relativně přesné čtení hodnot v mapě. Samotné zaplnění mapy interpolovanými hodnotami však není dostačující. Zvláště ceny pronájmů v rámci zájmového území České republiky často chybí a jsou omezeny převážně jen na velká města. Taktéž je pravděpodobné, že i v místech, na kterých se hodnoty vyskytují, jich není mnoho. K tomuto faktu lze dospět z identifikace červených, či fialových teček v mapě vycházejících z vysokých cen nemovitostí, které nebyly vyhlazeny interpolací s okolními hodnotami. Jednou z možností vzniku takovýchto fenoménů může být právě nedostatek vstupních hodnot, druhou pak neadekvátní nastavení interpolační metody. V době finalizace diplomové práce byla tato aplikace již nefunkční. Sumarizace kladů a záporů je uvedena v tabulce 1.

Tabulka 1: Shrnutí cenové mapy ze serveru bezreality.cz

Klady	Zápory
Barevná škála a legenda skládající se z více barevných tónů umožňující rychlé a relativně přesné čtení hodnot v mapě	Bez popisu zdrojových dat, metodiky zpracování dat a časového rámce
	Absence informace o počtu vstupních dat
	Bodové fenomény způsobené nevhodnou interpolací, malým počtem dat, či extrémy v datech

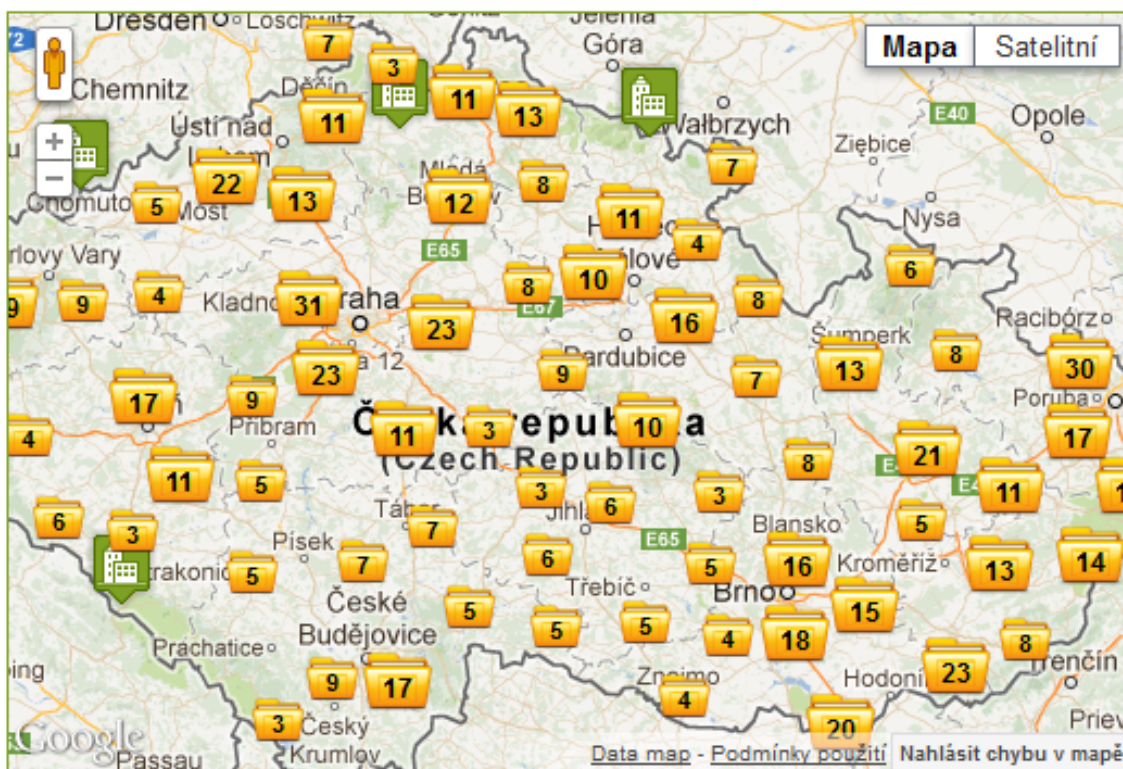


Obrázek 2: Cenová mapa Asociace realitních kanceláří

Jako další pak byla nalezena cenová mapa od Asociace realitních kancelářů. [7] Její vzhled je zobrazen na obrázku 2. U této mapy jsou již uvedené určité základní informace o zpracování vstupních dat. Dle údajů z webových stránek je mapa tvořena na základě informací o skutečně realizovaných cenách nemovitostí a dodavateli jsou členské kanceláře ARK ČR, nečlenské kanceláře a další subjekty. Lokalizace hodnot je tedy pravděpodobně realizována přímo na adresní body nemovitostí a tím pádem může dosahovat vyšší přesnosti než aplikace vyvíjená v této práci. S tímto problémem se však seznámíme v další kapitole. Informace o cenách jsou zde sbírány průběžně a přepočítání dat na zobrazované rastry probíhá měsíčně. Zda jsou staré hodnoty vymazány, či jsou stále zahrnuty do výpočtu průměru uvedeno není. Jestli jsou však hodnoty přesné a nejsou například nadsazeny, aby dosavadní nabídky vypadaly lukrativněji, není jasné. Na rozdíl od mapy ze serveru bezrealitky.cz je zde již možnost přepínání vrstev domů a bytů. U domů se jedná o prodej a u bytů jak o prodej, tak o pronájem. Pro velká měřítka aplikace využívá interpolaci se zohledněním okolí jednoho kilometru, pro měřítka malá pak zohledňuje okolí pěti kilometrů. Mírně nepřehledná je zde legenda. Jedná se o plynulý přechod dvou barevných tónů, a tím pádem je z prostého pohledu na určitou lokalitu poměrně obtížné určit relativně přesně zjišťovanou hodnotu. Cenová mapa však nabízí při kliknutí do mapy výpočet hodnoty z průměru čtverce o straně 500 metrů v daném místě při velkém měřítku. Při menších měřítcích se pak délka strany čtverce zvětšuje. Všechny tyto funkce a vlastnosti jsou dostupné zdarma, je tady však i možnost placeného vstupu, který však nebyl praktikován. Zároveň na stránkách chybí popis, či výčet zmíněných nadstandardních funkcí. Opět je uvedeno shrnutí v tabulce 2.

Tabulka 2: Shrnutí cenové mapy Asociace realitních kancelářů

Klady	Zápory
Popis zdrojových dat, metodiky zpracování dat a časového rámce	Absence informace o počtu vstupních dat
Skutečně realizované ceny nemovitostí	
Možnost získání přesného údaje o ceně kliknutím do mapy	Barevná škála a legenda neumožňující rychlé a relativně přesné čtení hodnot v mapě
Rozlišení údajů na domy a byty	



Obrázek 3: Cenová mapa Státního fondu rozvoje bydlení

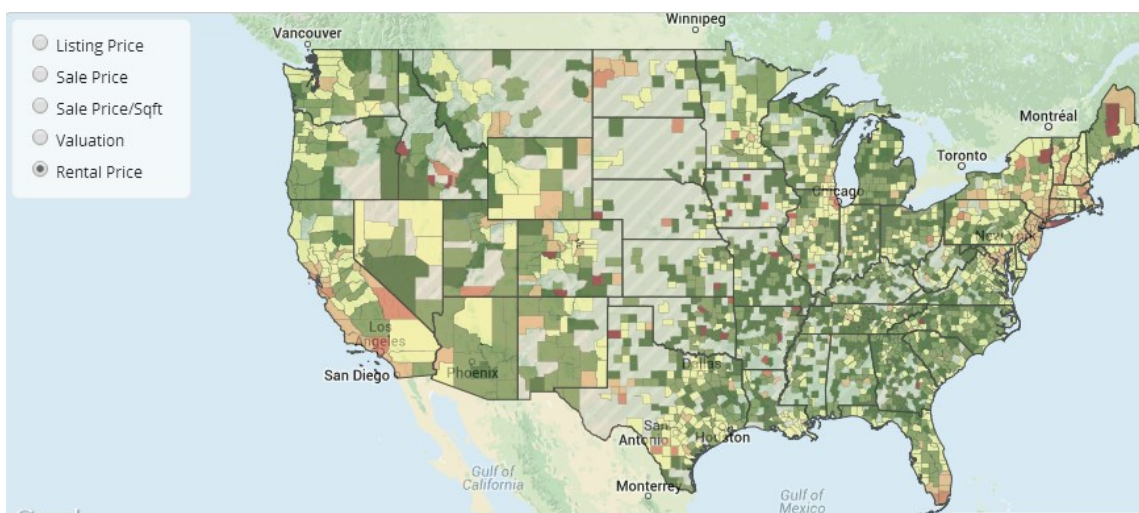
Poslední českou vyhledanou cenovou mapou je aplikace Státního fondu rozvoje bydlení. [8] Náhled je uveden na obrázku 3. V aplikaci jsou uvedeny pouze ceny nájemného ve vybraných obcích. V mapě jsou zavedeny údaje, které SFRB zcela převzalo od Ministerstva pro místní rozvoj ČR. Mapa nájemného vznikla z důvodu potřeby informovat nájemce a nájemníky o předpokládané výši budoucího nájemného po skončení účinnosti zákona o jednostranném zvyšování nájemného z bytu. Dále je na webových stránkách uvedeno, že se jedná pouze o informativní údaje sloužící jen jako orientační data. Orientační údaj, tedy odhad výše nájemného, zjištěný dle posudku znalce pro standardní byt, je zpracován pro určené obce, které měly více než 2000 obyvatel dle ČSÚ k 1. 10. 2010.

Oproti cenovým mapám realitních kanceláří zmíněných dříve, zde již nejsou použity rastrová data. Údaje dokonce v mapě nejsou zaznamenány vůbec. Při vyhledání zájmové obce musí uživatel kliknout na danou obec v mapě a následně si na stránkách SFRB pod mapovým oknem přečíst údaj o ceně stanovený znalcem. Chybí zde tedy rychlá a efektivní orientace v datech. V případě, že se v obci nachází lukrativní bytová zástavba, ale zároveň také ghetto, údaj zde uvedený znamená jen určitý průměr, ne příliš reprezentující realitu. Aplikace tedy nepracuje s vnitřní variabilitou v obci a úroveň ekologické chyby je v tomto případě velmi vysoká. Dalším nedostatkem pak může být datum aktualizace, u většiny zkoumaných obcí se jedná o rok 2011, u některých větších měst pak 2012. U této

cenové mapy, ale i u aplikací zmíněných dříve, je jako podkladová mapa využita mapa Google. Stejně jako cenová mapa ze serveru bezrealitky.cz byla tato aplikace v době finalizace diplomové práce nedostupná. Shrnutí kladů a záporů je opět uvedeno v tabulce 3.

Tabulka 3: Shrnutí cenové mapy Státního fondu rozvoje bydlení

Klady	Zápory
Popis zdrojových dat, metodiky zpracování dat a časového rámce	Prezentované ceny jsou pouhým odhadem
	Údaje nejsou zobrazeny adekvátní kartografickou metodou
	Pouze nájemní ceny bytů
	Možnost ekologické chyby v datech



Obrázek 4: National home prices page

Existuje možnost, že by cenové mapy na české scéně byly oproti těm zahraničním zaostalé, či jinak nedokonalé. Z toho důvodu by bylo vhodné pokusit se nalézt zahraniční ekvivalent cenových map zkoumaných u nás. Byla tedy nalezena cenová mapa ze serveru Trulia zahrnující zájmové území Spojených států amerických. [9] Náhled na tuto cenovou mapu je zprostředkován obrázkem 4. Server Trulia dle informací ze svých stránek je nestranným a nezávislým serverem zprostředkujícím vyhledávání a inzerce nemovitostí. Celkově se tedy jeví podobně jako náš server Sreality. Hlavním rozdílem oproti českým aplikacím je skutečnost, že ceny jsou zde prezentovány primárně za celý byt či dům. Celkově aplikace nabízí výběr ze zobrazení inzerčních cen, cen z uskutečněných prodejů, odhadních cen a cen nájmu. Pouze u cen uskutečněných prodejů je zde možnost zvolit přepočítání ceny na metr čtvereční. Dále zde není možnost filtrace mezi domy,

či byty. Z kartografického hlediska, ačkoliv jsou polygony v mapě vybarveny určitou barvou, legenda k mapě poskytována není a hodnotu ceny se uživatel dozví pouze najetím myši nad zkoumanou lokalitu. Ceny zde nejsou zobrazeny rastry z interpolovaných hodnot, ale ve vektorové formě prostřednictvím polygonů, kdy za daný polygon je vždy zobrazena mediánová hodnota. Počet hodnot, ze kterých je medián určen však znám není. V nejmenším měřítku je území členěno na okresy, ve větších oblastech stanovenými ZIP kódem a v největším měřítku pak samotnými bloky budov ve velkých městech. Celkově je daná aplikace srovnatelná s těmi českými. Pro tuto práci však nabídla rozumný příklad zobrazení cen pomocí kartogramů, který právě české aplikace neposkytly. Přehled kladů a záporů je uveden v tabulce 4.

Tabulka 4: Shrnutí cenové mapy ze serveru Trulia

Klady	Zápory
Zobrazení údajů pomocí kartogramů	Ve většině případů ceny nejsou přepočteny na m ²
	Absence informace o počtu vstupních dat
	Absence legendy
Možnost výběru typu ceny	Bez popisu zdrojových dat, metodiky zpracování dat a časového rámce
	Absence rozlišení na domy a byty

V době psaní této diplomové práce se pak na českém trhu objevil projekt cenové mapy transakčních cen. [10] Jedná se o webovou aplikaci vytvořenou na základě databáze transakčních cen, za které byly konkrétní nemovitosti skutečně prodány. Je zaměřena na část realitního trhu nových a zánovních bytů, tedy bytů postavených v nových bytových projektech od poloviny devadesátých let, kdy byla v České republice ukončena centrálně plánovaná výstavba panelových sídlišť, do současnosti. Mimo porovnávání nemovitostí dle jejich cen je jejím údajným klíčovým rysem porovnávání dle parametrů komfortu, služeb, lokace a technologie. Momentálně je však aplikace pouze přístupná přímými účastníky rezidenčního trhu. Laická veřejnost bude mít umožněn přístup teprve 1.1.2015. Projekt se však jeví ambiciózně.

4. Výběr vhodného realitního serveru

Před započítím jakékoliv činnosti bylo nutné vybrat vhodný realitní server. Takový server by měl být posuzován z více hledisek. Množství inzerátů, jejich kvalita, úplnost či přesnost jsou velmi důležité. Tato práce se však zabývá použitím RSS kanálu. Kromě kvality inzerátů je důležitá také kvalita samotného RSS kanálu. Pokud by byly jednotlivé inzeráty uloženy na serveru adekvátní, avšak nebyly prezentovány precizně právě v RSS kanálu, případně nebyly v něm prezentovány vůbec, stává se takový server nepoužitelným.

Mezi největší realitní servery u nás patří například sreality.cz, realitymix.cz, hyperreality.cz, hmreality.cz nebo reality.idnes.cz. Tyto a mnoho jiných serverů vyhledaných na internetu bylo tedy nutné projít a ověřit jejich vhodnost užití. Jediné sreality.cz poskytovaly adekvátní RSS kanál. Ostatní servery v naprosté většině případů neposkytovaly RSS kanál žádný, v méně případech pak velmi nevhodný. Příkladem nevhodného kanálu je například portál hmreality obsahující pouze informaci o typu objektu a odkazu na webovou stránku inzerátu.

4.1. Popis realitního serveru Sreality

Jedná se o nejnavštěvovanější český realitní server s průměrnou denní návštěvností přes 60 000 reálných uživatelů nabízející přes 190 000 nemovitostí od více než 2 700 realitních kanceláří i jednotlivců. Databáze realitního serveru shromažďuje nabídky nemovitostí z českého realitního trhu přímo od majitelů nemovitostí nebo realitních kanceláří. Zájemcům o koupi či pronájem nemovitosti zaručuje 24 hodin denně aktuální údaje a vyhledávání pomocí filtrů. Prezentace jednotlivých inzerátů obsahuje detailní popisy nemovitostí včetně fotografií, video prohlídek a mapy okolí. Zpoplatněním inzerátů a kontrolou obsahu dosahuje server maximálního zkvalitnění obsažených nabídek. [1]

Z výše uvedeného textu lze vyvodit několik důležitých informací. Navštěvovanost či oblíbenost serveru pro účel této práce není rozhodující. Důležitější je množství poskytovaných inzerátů, které nám zaručí kvalitní zaplnění databáze, či aktualizace 24 hodin denně, což umožní pružnější stahování inzerátů z RSS kanálu v reálném čase. Nejdůležitější informací je však zpoplatnění inzerátů a kontrola obsahu zaručující minimalizaci chyb a nepřesností v datech.

4.2. Popis RSS kanálu sreality.cz

Jelikož je RSS kanál jediným zdrojem informací pro vyvíjenou databázi, je nutné si jej podrobněji popsat za účelem lepší organizace a zpracování dat. V daném kanálu se kromě obecných informací jako je verze rss, názvu stránky, odkazu, popisu a loga vyskytuje pro tuto práci důležitých 10 položek s inzeráty. Při načtení kanálu se nám vždy zobrazí právě 10 položek. Pokud je kanál načten téměř okamžitě znovu, v době kdy na serveru nepřibyl žádný další inzerát, zobrazí se nám opět těchto 10 stejných inzerátů. Pokud je však kanál načten v době, kdy na server již nové inzeráty přibýly, budou některé, ne-li všechny, položky nové, právě dle počtu nových inzerátů. Pro příklad je uvedena jedna položka z kategorie prodeje bytů:

```
<item>
<title>Prodej, byt 2+kk, 48&nbsp;m2</title>
<link>
http://sreality.cz/detail/prodej/byt/2+kk/brno--trida-general-a-
piky/1009148764
</link>
<description>
Cena 1 590 000 Kč za nemovitost, třída Generála Píky, Brno, Panelová,
Osobní, 3. patro,
</description>
</item>
```

Samotná položka je oddělena dvojitou značkou <item>. V titulku se nachází informace, zda-li se jedná o prodej, či pronájem, velikost bytu z hlediska pokojů a výměra z hlediska užitné plochy. Bylo by tedy vhodné přesněji definovat užitnou plochu. Rozumí se jí plocha všech obytných a vedlejších místností a plocha příslušenství bytu, nezapočítává se plocha nebytových prostor. [11] Samozřejmě záleží na inzerentech, zda danou položku vyplní správným údajem, nebo ji zamění s plochou jinou. U prodeje či pronájmu bytů je výměra užitné plochy vhodná, problém však nastává při zpracování prodeje domů, o kterém se zmíníme později.

Následuje odkaz <link> na samotný inzerát. Odkaz na webovou stránku inzerátu bude v našem případě využit jako jednoznačný identifikátor daného inzerátu. V budoucnu by jej bylo možné využít k rozšíření této práce prostřednictvím vyhledávajícího bota. Takový bot by dostával z RSS kanálu informaci o nových inzerátech a procházel by stránky, na které je prostřednictvím RSS odkazován. Zde by mohl zjistit a uložit znatelně více informací, které se v samotném kanálu nevyskytují, případně by mohl do databáze stahovat i obrázky daných realit. Legálnost takového řešení by ovšem musela být ověřena.

Jako poslední položka je popis inzerátu skrytý ve značce <description>. Nachází se zde cena nemovitosti u prodeje, či měsíční nájem u pronájmu. Měsíční nájem je zde uváděn bez poplatků za vodu a energie. V případě částky 0 Kč, či

jiné velmi nízké hodnoty je vhodné takový inzerát odstranit, aby nedocházelo k znehodnocení dat při zpracování agregací, či interpolací. Filtrace dat bude popsána v kapitole 7.1.

V ideálním případě následuje ulice, na které se nemovitost nachází. Číslo popisné není uváděno, aby nedocházelo k obcházení realitních kanceláří při realizaci prodeje. Lokalizace inzerátu je tedy možná nejpřesněji na danou ulici. V případě relativně krátkých ulic nezasahujících do více městských částí je taková přesnost dostačující, v případě dlouhých ulic se jedná o problém, který bohužel není řešitelný automatizovaně. Jak již bylo zmíněno, název ulice jde v RSS kanálu uváděn pouze v ideálních případech. Jedná se o položku, která bývá v určité míře nevyplněna. Dle aktuální databáze (březen 2014) je inzerátů v prodeji bytů bez vyplněné ulice 14,1%, u jejich pronájmu 8,9% a u prodeje domů bohužel 59,7%. Absence této položky může být v lepším případě způsobena absencí názvů ulic v zájmových obcích, v horších případech nevyplněním při zadávání inzerátu. V případě agregace hodnot na kraje, okresy, či obce tato absence nehraje žádnou roli. Pokud však budeme chtít zobrazit hodnoty pro ulice v určité zájmové oblasti, tak již úbytek inzerce v dané agregaci bude znatelný, minimálně u prodeje domů. Řešit tento problém například přidělením průměrné hodnoty za danou obec všem ulicím v dané obci, by však přineslo výsledky zatížené ekologickou chybou.

Po problematickém názvu ulice následuje název obce, u kterého již potíže nenastávají. Poté je uveden typ zástavby. Ten je vybírán z domény hodnot Dřevěná, Cihlová, Kamenná, Montovaná, Panelová, Skeletová a Smíšená. Jako předposlední údaj v popisu inzerátu je uveden typ vlastnictví z domény hodnot Osobní, Družstevní a Jiné. Tento údaj bude mít při prodeji bytu významný vliv na jeho cenu. Na závěr je uvedeno patro ve kterém se byt nachází.

Jelikož je obsah RSS kanálu při prodeji domů mírně odlišný, uvedeme taktéž příklad pro něj:

```
<item>
<title>Prodej, dům rodinný, 140 m2</title>
<link>
http://sreality.cz/detail/prodej/dum/rodinny/praha-hostivar-herbenova/1844958044
</link>
<description>
Cena 6 100 000 Kč za nemovitost, Herbenova, Praha, Samostatný , byty 0,
</description>
</item>
```

Avizované rozdíly se nachází již v samotném titulku. Není zde uveden počet pokojů. Místo toho je uvedeno o jaký typ domu se jedná z domény dům rodinný, činžovní, vila, na klíč, dřevostavba a nízkoenergetický. Stejně jako u bytu je

uvedena užitná plocha. Při koupi domu by však mohly zájemce zajímat i jiné výměry, například plocha zastavěná, či plocha pozemku. Tyto údaje zde bohužel uvedeny nejsou.

V popisu inzerátu jsou uvedeny na rozdíl od bytů údaje o poloze domů z domény Řadový, Rohový, V bloku a Samostatný. Jako poslední je pak informace o počtu bytů v daném domě. Tuto položku inzerenti často nevyplňují a jako implicitní hodnota zde bývá vyplněno 0 bytů. Položky typ zástavby, vlastnictví a patro uvedené u bytů se zde logicky nenacházejí.

4.3. Žádost o spolupráci s portálem seznam.cz

Ačkoliv se může zprvu zdát, že užití RSS kanálu je naprosto legálním zdrojem dat, po prostudování autorského práva v kartografii a geoinformatice situace tak jasná již není. [12] Podmínky užití kartografických děl a GIS produktů jsou poměrně přísné a tvorba jakéhokoliv díla, které vzniklo úpravou nebo zpracováním díla jiného bez souhlasu jeho autora, je porušením autorského zákona. Týká se to například i digitalizace určitých prvků z leteckého snímku nebo z jiné mapy. U digitálních produktů GIS a u softwaru se autorsko-právní ochrana řídí především konkrétními legislativními opatřeními, která zahrnují například ochranu databází a počítačových programů.

Padlo tedy rozhodnutí nejdříve kontaktovat odpovědné osoby ze společnosti Seznam a požádat je o spolupráci minimálně v tom rozsahu, aby bylo dosaženo písemného souhlasu s užitím dat získaných prostřednictvím RSS kanálu. Byl tedy vypracován stručný a věcný dokument popisující účel a zaměření této diplomové práce spolu s předběžnými vizualizacemi možných mapových výstupů a tento dokument byl vedoucím diplomové práce jménem Institutu Geoinformatiky zaslán společnosti Seznam.

V následné telefonické a e-mailové komunikaci bylo potvrzeno, že data z RSS kanálu jsou volně dostupná a jejich užití je tedy legální. Samozřejmě bylo požadováno i písemné vyjádření. Na tom údajně pracovalo právníké oddělení společnosti Seznam a dosud nebylo doručeno. Kromě telefonického obeznámení s volností užití daných dat, však je k dispozici i záznam z e-mailové komunikace, ve které je souhlas uveden taktéž. Po dokončení diplomové práce bude společnosti Seznam zaslána její elektronická verze a odkaz na webovou aplikaci. Může se jednat o inspiraci, kterou by mohli využít a následně implementovat vlastní verzi aplikace, umožňující koncovým uživatelům jejich portálu lepší orientaci v inzertních cenách.

5. Návrh a realizace databáze

Pro možnost následného zpracování a využití dat z RSS kanálu je potřeba data ukládat do předpřipravené databáze. Vzhledem k periodické aktualizaci RSS kanálu zmíněné dříve, je vhodné mít databázi umístěnou na stroji, který je připojen k internetu neustále. Z toho důvodu byl zřízen přístup na školní server studgisweb. Na tomto serveru je pro potřeby studentů předinstalovaná sada svobodného softwaru jménem LAMP. Tato sada je užívaná jako platforma pro implementaci dynamických webových stránek a zahrnuje tyto technologie:

- Linux – operační systém
- Apache – webový server
- MySQL – databázový systém
- PHP – skriptovací programovací jazyk

Server je tedy primárně určen pro realizování webových prezentací studenty. V době započetí činností na diplomové práci však nebylo dostupné komplexní serverové řešení umožňující jak realizaci databáze, stahování dat a následnou webovou prezentaci prostorových dat. Pro potřeby realizace databáze a stahování informací z RSS kanálu byl však tento server dostačující. Každá ze součástí LAMP je v této práci více, či méně využita.

K realizaci databáze byl tedy využit relační databázový systém MySQL. Jedná se o systém vytvořený švédskou firmou MySQL AB, nyní vlastněný společností Sun Microsystems, dceřinou společností Oracle Corporation. Je považován za úspěšného průkopníka dvojího licencování. Je totiž dostupný jak pod bezplatnou licenci GPL, tak pod komerční placenou verzí. V této práci je užitá licence bezplatná. Jedná se o multiplatformní databázi schopnou fungovat na různých operačních systémech. Komunikace s ní probíhá, jak již název napovídá pomocí jazyka SQL. Stejně jak u ostatních SQL databází, se jedná o určitý dialekt tohoto jazyka s některými rozšířeními. Pro snadnou implementovatelnost, výkon a volné šíření má vysoký podíl na v současné době používaných databázích.

K databázovému systému, je možné se připojit pomocí emulátoru terminálu PuTTY. Jedná se o klient protokolů Secure Shell, Telnet, rlogin a TCP. Pracovat s rozsáhlou databází, kontrolovat data a filtrovat chyby, by však nejspíše bylo pomocí terminálového připojení poměrně neobratné. Z toho důvodu existuje programový systém phpMyAdmin. Je to nástroj napsaný v jazyce PHP umožňující jednoduchou správu obsahu databáze MySQL prostřednictvím webového rozhraní. V současné době umožňuje vytvářet / rušit databáze, vytvářet / upravovat / rušit tabulky, provádět SQL příkazy a spravovat klíče. Zároveň podporuje export a import dat v mnoha formátech. [13] Tento systém byl taktéž

na serveru již nainstalován a pro potřeby této práce byl hojně využíván. Zjištěnou nevýhodou je omezení maximální velikosti importovaných dat na 10MB, v tomto případě bylo nutné využít připojení k databázi přes terminál. Jinou možností řešení je rozdělit vstupní soubor na několik částí a ty následně importovat samostatně.

5.1. Tabulky pro uložení stažených dat

Data se v rámci databáze ukládají do tabulek. Vzhledem k povaze stahovaných dat, by data mohly být uloženy do dvou tabulek, tedy byty a domy. Nicméně z důvodu lepší přehlednosti, organizovatelnosti a následného filtrování byl zvolen počet tří tabulek. Byly tedy vytvořeny zvláště tabulky pro prodej bytů, pronájem bytů a prodej domů.

Vzhledem k určité strohosti RSS kanálu nebylo nutné se zabývat zestručněním jeho obsahu. U prodeje a pronájmu bytů by bylo jistě přípustné v daném záběru diplomové práce nevytvářet v tabulkách sloupce pro typ budovy, či patro, ve kterém se daný byt nachází. Tyto položky totiž nejsou v dané práci nijak využívány. Pokud by takových podobných nevyužitelných položek bylo více, určitě by bylo jejich zahrnutí v tabulkách zvažováno. V pouhém počtu dvou nadbytečných sloupců však byly uchovány pro případné využití v budoucnu.

U prodeje a pronájmu bytů byly tedy vytvořeny sloupce *id*, *datum*, *title*, *url*, *typ*, *typ_nemovitosti*, *plocha*, *cena*, *cena_za_1m*, *ulice*, *mesto*, *budova*, *vlastnictvi* a *patro*. Sloupec *id* slouží jako jednoznačný umělý identifikátor v rámci tabulky. *Datum* udává termín uložení inzerátu do databáze s přesností na dny a sloupec *title* reprezentuje řetězec titulu RSS kanálu před rozdělením na jednotlivé části. Význam zbylých sloupců byl vysvětlen v kapitole 4.2.

id	datum	typ_nemovitosti	plocha	cena	cena_za_1m	ulice	mesto	budova	vlastnictvi	patro
146935	11/03/2014	byt 1+1	44	1300000	29545.455	Husitská	Olomouc	Cihlová	Osobní	0. patro
146934	11/03/2014	byt 1+1	44	330000	7500.000	Fibichova	Jáchymov	Cihlová	Osobní	1. patro
146933	11/03/2014	byt 4+1	84	2250000	26785.714	Prušánecká	Brno	Panelová	Družstevní	4. patro
146932	11/03/2014	byt 2+kk	64	1850000	28906.250	K Čínovu	Klatovy	Cihlová	Osobní	1. patro
146931	11/03/2014	byt 3+kk	85	2290000	26941.176	Valdenská	Olomouc	Cihlová	Osobní	2. patro
146930	11/03/2014	byt 3+kk	85	1900000	22352.941	Valdenská	Olomouc	Cihlová	Osobní	1. patro
146929	11/03/2014	byt 4+kk	91	6637000	72934.066	Chlupáčova	Praha	Cihlová	Osobní	3. patro

Obrázek 5: Náhled tabulky prodeje bytů

Na obrázku 5 je pro představu zobrazen náhled na několik záznamů tabulky prodeje bytů prostřednictvím systému phpMyAdmin. Z důvodu velikosti obrázku

nebyly vybrány sloupce *title*, *typ* a *url*. U položky *id* byl nastaven datový typ *int* pro běžně velká celá čísla a speciální funkčnost *auto increment* pro generování unikátního čísla při vložení nového záznamu do tabulky. *Datum*, *url*, *typ*, *typ_nemovitosti*, *ulice*, *mesto*, *budova*, *vlastnictvi* a *patro* mají přidělený datový typ *varchar*. Výhodou tohoto datového typu je skutečnost, že pokud je vložený řetězec kratší, než nastavíme, chybějící znaky se nedoplňují mezerami, což například u názvů ulic či měst výrazně snižuje velikost tabulky a tedy objem databáze. Cena má datový typ *int*, *cena_za_1m* datový typ *double*, tedy číslo s desetinnou čárkou. U všech textových polí je nastaveno kódování *UTF-8_czech*, aby byla zachována správná diakritika. Taktéž je vhodné nastavení indexování sloupců *mesto* a *ulice*, zrychlí se tak vyhledávací a dotazovací procesy v databázi, zvláště pokud bude dotaz obsahovat určitou variantu spojování tabulek.

Stejně jako tabulky prodeje a pronájmů bytů, byla vytvořena i tabulka prodeje domů. S tím rozdílem, že neobsahuje sloupce *budova*, *vlastnictvi* a *patro*. Místo nich jsou zde sloupce *zastavba* a *jednotek*. Jejich význam je opět uveden v kapitole 4.2. Situace ohledně datových typů, kódování a indexování je stejná jako u prodeje a pronájmu bytů.

5.2. Podpůrné tabulky pro prostorovou reprezentaci

Tabulky z předchozí kapitoly jsou určeny k uložení dat stažených z RSS kanálu. Jedná se však o surová data, která je nutno před připojením k určité geometrické složce předzpracovat. Předzpracováním je myšlena například agregace dat k určitým územním celkům jako kraje, okresy a obce, nebo k samotným ulicím, kterých se inzeráty týkají. Taktéž byla testována možnost připojení X a Y souřadnic k jednotlivým inzerátům, aby byla umožněna interpolace inzerčních cen. Dostupné GIS produkty použitelné pro následné zpracování dat nicméně takové předzpracování nepodporují. Používaný produkt ESRI ArcGIS podporuje spojování tabulek v poměru 1:1, či 1:N, nepodporuje však agregační funkce. Pro případ připojení souřadnic k inzerátům je pak potřeba spojení na základě dvou sloupců, což také použitý geoinformační systém nepodporuje. Pomocí podpůrných tabulek a jazyka SQL lze však v prostředí databáze tyto úkony jednoduše provést.

KOD_OBEC_P	NAZ_OBEC_P	KOD_ORP	NAZ_ORP	LAU1_P	NAZ_LAU1_P	CZNUTS3	NAZ_CNUTS3
500011	Želechovice nad Dřevnicí	7213	Zlín	CZ0724	Zlín	CZ072	Zlínský kraj
500020	Petrov nad Desnou	7111	Šumperk	CZ0715	Šumperk	CZ071	Olomoucký kraj
500259	Veřovice	8105	Frenštát pod Radhoštěm	CZ0804	Nový Jičín	CZ080	Moravskoslezský kraj
500291	Vřesina	8119	Ostrava	CZ0806	Ostrava-město	CZ080	Moravskoslezský kraj
500496	Olomouc	7107	Olomouc	CZ0712	Olomouc	CZ071	Olomoucký kraj
500526	Bělkovice-Laštany	7107	Olomouc	CZ0712	Olomouc	CZ071	Olomoucký kraj
500623	Bílá Lhota	7105	Litovel	CZ0712	Olomouc	CZ071	Olomoucký kraj
500801	Blatec	7107	Olomouc	CZ0712	Olomouc	CZ071	Olomoucký kraj

Obrázek 6: Náhled tabulky obcí

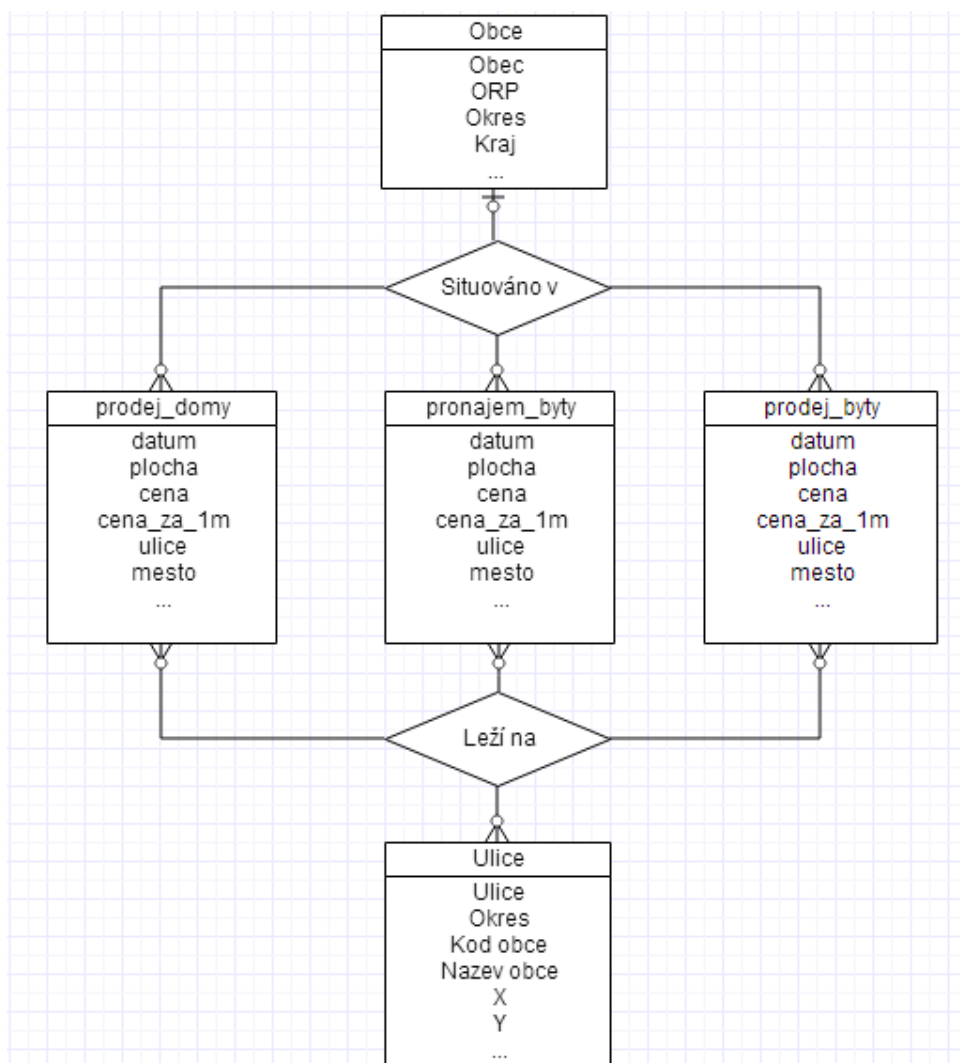
Jako první byla tedy vytvořena tabulka *Obce*. Náhled na několik prvních řádků je zobrazen na obrázku 6. Celkem má daná tabulka 6351 řádků, kdy každý řádek reprezentuje jednu unikátní obec. Obsah tabulky je převzat z vrstvy obcí České republiky, poskytnuté Registrem sčítacích obvodů. Daná vrstva bude taktéž v této práci později využita. Při následném spojení výstupní tabulky z databáze s touto vrstvou v prostředí GIS by nemělo dojít k nesrovnalostem způsobenými malými rozdíly v názvech krajů či okresů (např. „Vysočina“ vs „kraj Vysočina“ atd.). V tabulce se vedle názvu samotné obce nachází název obce s rozšířenou působností, název okresu a název kraje, do kterých daná obec spadá. Mimo to je zde ještě umělý identifikátor, kódy obcí, ORP, okresů a krajů.

FID	NAZEV_UL_A	LAU1	KOD_OBCE	NAZ_OBCE	X	Y
0	Brněnská	CZ0100	554782	Praha	464838,0291	5541222,471
1	U břehu	CZ0100	554782	Praha	467394,6727	5543182,983
2	U Prioru	CZ0100	554782	Praha	451027,8438	5548488,847
3	Hrazanská	CZ0100	554782	Praha	460028,0653	5536938,235
4	Vinořská	CZ0100	554782	Praha	469640,4278	5554057,482
5	K Brnkám	CZ0100	554782	Praha	459050,6218	5555752,874

Obrázek 7: Náhled tabulky ulic

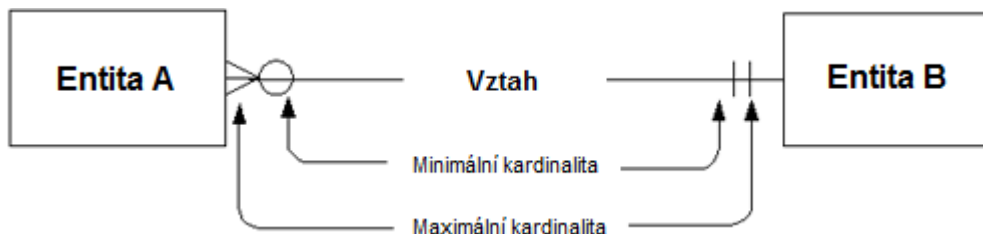
Jako poslední byla vytvořena tabulka ulic jejíž náhled je na obrázku 7. Celkem je v dané tabulce 170100 řádků. Tabulka je vytvořena z bodové vrstvy uliční sítě, taktéž poskytnuté Registrem sčítacích obvodů. Pokud se jedná o ulici krátkou, je v tabulce prezentována jen jedním řádkem, tedy jedním párem souřadnic. Pokud se však jedná o ulici delší, je v tabulce reprezentována více řádky, a tedy více páry souřadnic. Vzdálenosti mezi jednotlivými body jedné ulice jsou zhruba 100 metrů. V případě větší členitosti linie ulice jsou tyto vzdálenosti menší. Dané body jsou pravděpodobně středy úseků liniové vrstvy ulic, ze kterých byla tato bodová vrstva vygenerována. V tabulce je uveden opět umělý identifikátor, název ulice, kód okresu, název obce a souřadnice X,Y v souřadném systému UTM Zone 33N.

5.3. Struktura databáze



Obrázek 8: Entitně relační model databáze

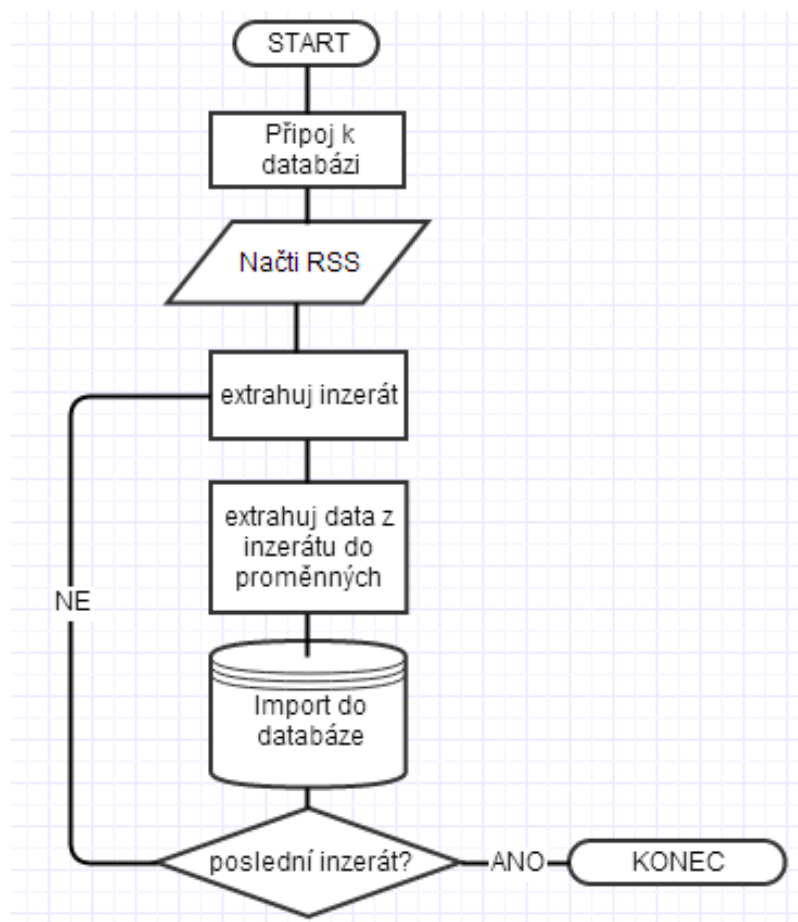
Pro detailní zobrazení vztahů mezi jednotlivými tabulkami v databázi obvykle slouží entitně relační model. Takový model, zjednodušený o několik atributů a s upravenými názvy sloupců, je zobrazen na obrázku 8. Vztahy v něm jsou vykresleny metodou Crow's foot (vraní noha), pro přehlednost je uvedeno vysvětlení na obrázku 9. Kruh znamená 0, svíslá čára 1, noha znamená mnoho.



Obrázek 9: Crow's foot data model

6. Inovace postupu pro stahování a import dat do databáze

Ve fázi, kdy již byla databáze připravena pro import dat, bylo nutné aplikovat určitý způsob čtení, zpracování a rozboru RSS kanálu s následným vložení výsledků do databáze. Čtení a uložení RSS kanálu obvykle obstarává RSS čtečka. Volně stažitelných čteček existuje mnoho, neposkytují však možnosti rozboru kanálu a importu dat do databáze. Zároveň by daná čtečka měla být umístěna na stejném stroji, na kterém se nachází databáze. Z toho plyne požadavek na funkčnost v systému Linux. Vhodným řešením tedy bylo vyvinout skript v některém z programovacích jazyků podporovaných využitým serverem. V předchozí kapitole bylo zmíněno předinstalování sady svobodného softwaru LAMP, jehož součástí je programovací a skriptovací jazyk PHP. Jazyk PHP je určen především pro programování dynamických internetových stránek a webových aplikací například ve formátu HTML, XHTML, či WML. Je však možné jej využít i k tvorbě konzolových aplikací. Interpret PHP skriptu je pak možné vyvolat pomocí konzole, v tomto případě PuTTY. Funkcionalita požadovaného skriptu je v hrubé formě vyobrazena na vývojovém diagramu z obrázku 10.



Obrázek 10: Funkcionalita stahovacího skriptu

Na tvorbě obdobného typu skriptu pracoval již v roce 2011 student Stanislav Pučan. Vypracoval v něm připojení k databázi, čtení RSS dokumentu, jeho základní rozklad dle značek a následný import do databáze. Poté však práce zanechal. Tento skript trpěl nedokonalostmi jako vytváření duplicit, či ukládání položek na špatné pozice v rámci záznamu v databázi. Navzdory nedostatkům, byl využit jako základ pro tvorbu skriptu vlastního, upraven a vylepšen.

6.1. Kód skriptu

Kód skriptu se skládá z několika částí. Při vysvětlování procesu stahování a zpracování budou pro lepší orientaci v tomto textu uvedeny fragmenty kódu. První část je deklarace přístupových údajů k databázi a samotné připojení. Následuje deklarace a definice funkce zprostředkující zpracování RSS kanálu. Tato funkce pak obstarává veškerou činnost od načtení kanálu, přes jeho rozbor a zpracování, až po samotný import položky do databáze. Rozbor kanálu spočívá nejdříve v oddělení jednotlivých značek každé položky v kanálu.

```
foreach ($rss->getElementsByTagName('item') as $node) {
$title = $node->getElementsByTagName('title')->item(0)->nodeValue;
$url    = $node->getElementsByTagName('link')->item(0)->nodeValue;
$desc   = $node->getElementsByTagName('description')->item(0)->nodeValue;
```

Jak již bylo zmíněno v kapitole 4.2, v samotné značce title a description je více informací, takže je nutné aplikovat rozbor na základě určitých kritérií i v rámci těchto značek. Titulek se skládá vždy ze tří položek oddělenými čárkami, je tedy možné jej považovat za určitou řadu a pomocí indexování těchto položek vybrat zvláště položky typ, typ nemovitosti a plochu. U plochy je následně nutné oddělit od hodnoty výměry jednotku m².

```
$array_title = explode(",", $title);
$typ = trim($array_title[0]);
$typ_nemovitosti = trim($array_title[1]);
$plocha = trim($array_title[2]);
$plocha = preg_replace('/^[^0-9]/', '', $plocha);
```

V případě, že by byl obsah značky description vždy plný, tak jak je tomu u title, řešil by se rozbor stejně, bohužel tomu tak není. Jak již bylo dříve popsáno, v inzerci chybí často název ulice. Pro potřeby indexování tedy neznáme přesný počet položek v řadě. Ulice je však v dané značce na druhém místě. Odebereme-li tedy cenu nemovitosti a následně procházíme řadu od konce, na poslední pozici, respektive první položce dané řady, nám pak zbyde buď prázdná položka, nebo hledaný název ulice. K procházení řady od konce a ukládání jednotlivých položek slouží funkce array_pop.

```
preg_match("/Cena (.*) Kč za nemovitost/", $desc, $cena);
    $desc = trim(str_ireplace ($cena[0], "", $desc), " ");
    $cena = preg_replace('/[^0-9]/', '', $cena[1]);
$array_desc = explode ("", $desc);
    $patro = trim(array_pop($array_desc));
    $vlastnictvi = trim(array_pop($array_desc));
    $budova = trim(array_pop($array_desc));
    $mesto = trim(array_pop($array_desc));
    $ulice = trim(array_pop($array_desc));
```

Taktéž byla již dříve zmíněna absence položky typ zástavby u prodeje domů. Obsah této položky vychází z domény „Samostatný, Řadový, Rohový a V bloku“. V případě přidávání hodnot proměnným, jak tomu bylo v předchozím případě, by se při absenci typu zástavby v daném inzerátu hodnoty uložily do špatných proměnných – proměnná typ zástavby by obsahovala město, město by obsahovalo ulici a na ulici by již v inzerátu nezbyl záznam. Tento problém byl řešen jednoduše podmíněným příkazem if. Byla porovnána hodnota druhé položky od konce řady se seznamem typů zástavby. V případě shody se záznamy uloží normálně, v případě neshody je položce typ zástavby přidána nulová hodnota a ostatním jsou přidány hodnoty správné.

Před importem do databáze je již nutné pouze určit aktuální datum a vypočítat cenu za metr čtvereční. Datum je určen pomocí php funkce Time(), cena je vypočtena vydělením hodnot proměnných ceny a plochy.

Nyní by položka z RSS kanálu mohla být nahrána do připojené databáze. V načteném kanálu však nemusí být žádné nové položky. Častým jevem inzerce nemovitostí je taktéž změna ceny, přesněji pak postupné snižování ceny. V případě přímého nahrávání, by se v databázi hromadily duplicitní záznamy, dle výše zmíněných problémů, buď se stejnou cenou, nebo s cenou změněnou. Toto lze vyřešit porovnáním adresy url z kanálu se všemi adresami v databázi. V případě, že se daná url adresa v databázi nenachází, je záznam standardně vložen, v opačném případě je aktualizována pouze cena a cena za metr čtvereční. Kontrolování zda se cena liší nebylo praktikováno, pouze by to zpomalovalo provedení dotazu, a tak je cena aktualizována bez ohledu na to, jestli je jiná nebo není.

```
$q = mysql_query ("SELECT id,url FROM prodej_byty_cr WHERE url='$url'
LIMIT 1");
if (mysql_num_rows($q) == 0) {
    mysql_query("INSERT INTO prodej_byty_cr (title, url, typ,
    typ_nemovitosti, plocha, cena, cena_za_lm, ulice, mesto, datum,
    budova, vlastnictvi, patro)
    VALUES ('$title', '$url', '$typ', '$typ_nemovitosti', '$plocha',
    '$cena', '$cenazalm', '$ulice', '$mesto', '$datum', '$budova',
    '$vlastnictvi', '$patro)");
```

```
} else {  
    if($z=mysql_fetch_array($q)) {  
        $q2="UPDATE prodej_byty_cr set cena=".$cena.",  
        cena_za_lm=".$cena_za_lm." where id=".$z['id'];  
        mysql_query($q2);  
    }  
}
```

Touto částí php skript pro stahování a import dat končí. Plná verze všech tří skriptů (zvláště pro prodej a pronájem bytů a prodej domů) pro případné prostudování či přepracování je uložena na CD k této práci přiloženém.

6.2. Automatizace stahování

Samotný skript lze spustit pomocí dříve zmíněné emulace konzole PuTTY. Poté jednorázově proběhne pro daný RSS kanál. Takové jednorázové spuštění by však bylo výrazně neefektivní a mnoho záznamů do databáze by nepřineslo. Proto je potřeba využít jakýsi automatizovaný způsob. Nejjednodušším řešením je využití aplikace Cron. Jedná se prakticky o obdobu plánovače úloh systému Windows. Zadání úloh se zde uskutečňuje textovou formou ve specifickém souboru, který je otevřen po napsání „crontab -e“ do konzole. Před interpretem a cestou k souboru, který má být spuštěn se vyskytuje 5 pozic určujících čas spuštění. Ty postupně určují minutu, hodinu, den, měsíc a den v týdnu. V případě hvězdičky je pole ignorováno a příkaz se provede pokaždé. Vzhledem k aktualizaci RSS kanálu každou minutu vypadají námi naplánované úlohy takto:

```
* * * * * php /home/mat740/prodej_byty_cr.php  
* * * * * php /home/mat740/pronajem_byty_cr.php  
* * * * * php /home/mat740/prodej_domy_cr.php
```

7. Filtrace, agregace a export dat

Po dostatečně dlouhé době periodického stahování dat z RSS kanálu do prostředí databáze je možné přejít k samotnému zpracování. V této práci byly data do prostředí databáze stahována po celý rok 2013, pro prodej a pronájem bytů pak i poslední dva měsíce roku 2012. Kratší časové řezy by z hlediska zaplnění území České republiky inzertními nabídkami nemusely být dostatečné. Zároveň trh s nemovitostmi není natolik dynamický, aby byly kratší časové řezy vyžadovány. První kroky zpracování stažených dat jsou realizovány již v prostředí databáze. Jedná se o filtraci, agregaci a export ve vhodném formátu pro následné zpracování v prostředí GIS. Před realizací těchto kroků je nutné přerušit import nových inzerátů do databáze. Nejsnadněji to lze provést smazáním obsahu souboru crontab. V případě nepřerušení stahování inzerce by docházelo u filtrace ke zpětnému naplňování databáze chybnými záznamy a u agregace ke konfliktům během řešení SQL dotazu.

7.1. Filtrace dat

Jak již bylo zmíněno, obsah inzerce nemovitostí realitního serveru s reality je zpoplatněn a kontrolován, což by mělo přispět k minimalizaci chyb a překlepů. Nicméně se chyby v inzerci přesto vyskytují. Jedná se buď o neúmyslné překlepy, nebo například úmyslné nezadání některých údajů. V každém případě je nutné vliv těchto chyb minimalizovat vhodnou filtrací.

Překlepy, či špatné formulace v názvech měst a ulic nemohou být ovlivněny z hlediska samotných tabulek pro uložení dat. Při spojení s pomocnými tabulkami obcí a ulic však chybně zadané názvy nenajdou svůj protějšek, a tedy budou ze zpracování vynechány. Objem chybně zadaných názvů je však velmi nízký, řádově se pohybuje do 1%.

Závažnější situace nastává u údajů, které jsou pro tuto práci kritické. Jedná se o samotné ceny nemovitostí a jejich plochy. Pokud by do agregáčnických funkcí vstupovaly chybné hodnoty, docházelo by ke zkreslení finálních výsledků. Zvláště pak v oblastech s menší hustotou inzerce by tyto chybné hodnoty mohly viditelně ovlivnit výslednou cenovou mapu.

Výhodou při zpracování podmínek filtrace je skutečnost, že inzeráty za poslední měsíc či dva lze stále vyhledat dle uložených url adres na samotném realitním portálu. Tuto možnost je vhodné využít právě při stanovení krajních hodnot, odkdy by nemovitosti s přehnaně nízkou, či vysokou hodnotou ceny nebo plochy měly být vyfiltrovány ze zpracování.

Jako první, nejčtenější a nejvíce žřetelné chyby byly identifikovány absence cen. Při nezadání této položky je v RSS kanálu a následně v databázi evidována cena za nemovitost 0 Kč. Taktéž se ve značně menším počtu ve všech tabulkách vyskytovaly jiné velmi nízké ceny. K eliminaci těchto očividných chyb byla tedy stanovena skupina hodnot eliminující nepravděpodobné ceny pro všechny tabulky. U hodnoty pronájmu se jedná o 500 Kč a u prodeje o 50 000 Kč za nemovitost.

```
START TRANSACTION;
DELETE FROM pronajem_byty_cr WHERE cena <500;
DELETE FROM prodej_byty_cr WHERE cena <50000;
DELETE FROM prodej_domy_cr WHERE cena <50000;
COMMIT;
```

Podobný problém nastával i u položky plocha, která sice byla inzerujícími uživateli realitního portálu vyplňována nepoměrně častěji, nicméně i zde se vyskytovaly hodnoty nulové, či velmi malé. Častým zvykem bylo například místo výměry užité plochy domu či bytu zadávat do této položky počet místností dané nemovitosti. Zde je již problém, že mohou existovat byty velmi malé a při stanovení hodnoty filtru musíme být opatrní. Například byty z domény „garsonka“, „1+kk“ či „byt jiný“ mohou disponovat výměrou od 7 m². Byly tedy vyfiltrovány byty s výměrou pod 7 m², domy s výměrou pod 30 m² a byty z domén jiných než výše zmíněných s výměrou pod 15 m².

```
START TRANSACTION;
DELETE FROM `pronajem_byty_cr` WHERE plocha < 7;
DELETE FROM `prodej_byty_cr` WHERE plocha < 7;
DELETE FROM `prodej_domy_cr` WHERE plocha < 30;
DELETE FROM `pronajem_byty_cr` WHERE plocha < 15 AND typ_nemovitosti NOT
IN ("byt garsonka", "byt jiný", "byt 1+kk");
DELETE FROM `prodej_byty_cr` WHERE plocha < 15 AND typ_nemovitosti NOT IN
("byt garsonka", "byt jiný", "byt 1+kk");
COMMIT;
```

Zároveň nesmíme zapomínat, že problém chyb v cenách nemusí být orientovaný pouze k hodnotám nízkým, ale taktéž k hodnotám vysokým. Je tedy potřeba stanovit limity, za kterých je možné cenu dané nemovitosti považovat za nereálnou. V nabídce se však vyskytují velké činžovní domy, luxusní vily, či byty v historických centrech měst, jejichž hodnoty extrémně vysoké být mohou. Ve všech případech je však jejich vysoká cena spojována s relativně velkou užžitnou plochou. Dotaz byl tedy stanoven následovně.

```
START TRANSACTION;
DELETE FROM pronajem_byty_cr WHERE cena > 80000 AND plocha < 100;
DELETE FROM prodej_byty_cr WHERE cena > 50000000 AND plocha < 100;
```

```
DELETE FROM prodej_domy_cr WHERE cena > 50000000 AND plocha < 100;  
COMMIT;
```

V inzerci se taktéž nacházelo mnoho nemovitostí s relevantními hodnotami plochy a ceny, jejichž poměr však vykazoval hodnoty velmi nepravděpodobné. Z toho důvodu byly inzeráty s cenami za metr čtvereční u prodeje bytů menší než 1000 Kč a inzeráty domů s cenami za metr čtvereční menšími než 500 Kč rovněž odstraněny.

V závěru filtrace byly řešeny specifické problémy, kdy například u prodeje domů inzerenti pravděpodobně nezadali plochu užitnou, ale plochu pozemku, tím pádem cena za metr čtvereční byla výrazně nízká. Za povšimnutí taktéž stojí filtrace prodeje bytů osobního vlastnictví s cenou pod 150 tisíc. Zatímco při založení družstva by taková cena nového bytu mohla být reálná, u osobního vlastnictví při určité velikosti bytu nikoliv.

```
START TRANSACTION;  
DELETE FROM `prodej_domy_cr` WHERE cena_za_lm < 1000 AND plocha > 500;  
DELETE FROM prodej_byty_cr WHERE cena < 150000 AND vlastnictvi = "Osobní"  
AND plocha > 60;  
COMMIT;
```

7.2. Agregace a výběr dat pro export

Po úspěšné filtraci inzerce od chybných záznamů bylo přistoupeno k agregaci dat. Agregací je myšleno užití určité statistické funkce, pomocí které systém řízení báze dat umožňuje seskupit vybrané řádky dotazu a spočítat nad nimi výsledek vybrané aritmetické nebo statistické funkce.

Agregace je nezbytná z důvodu potřeby připojení inzertních nabídek k určité prostorové datové složce. Toto připojení je prováděno v prostředí některé z dostupných GIS aplikací. V tomto případě využívaný produkt ESRI ArcGIS nepodporuje agregaci tabulkových dat před jejich připojením k souboru shapefile, a tedy je nutné agregaci provést již v databázi.

V této fázi práce ještě nebylo plně rozhodnuto jakým způsobem budou inzertní data ve finální aplikaci prezentovány. Existují možnosti vektorového zobrazení dat formou kartogramů určitých samosprávních celků a ohodnocených linií ulic, či zobrazení pomocí rastrů interpolovaných hodnot cen nemovitostí. Aby bylo možné určit kterým směrem se bude aplikace ubírat, je nutné otestovat obě možnosti. Je nutné tedy provést jak agregaci dat na samosprávní celky a ulice, tak výběr vhodný pro následnou interpolaci, u kterého se už ovšem nebude jednat o agregaci.

Za účelem vytvoření kartogramů samosprávních celků je nutné seskupit data z tabulek pronájmu bytů a prodeje bytů i domů vždy dle obce, okresu a kraje. Taktéž by bylo možné aplikovat seskupení na obce s rozšířenou působností. Vzhledem k poměrně velkým rozdílům rozloh jednotlivých obcí s rozšířenou působností však nebylo toto seskupení realizováno. Pro každou tabulku tedy byla sestavena série SQL dotazů uskutečňující potřebné seskupení. Při seskupení na název obce je potřebná pouze tabulka inzertních nabídek. U agregací nabídek do krajů a okresů je již nutno využít pomocnou tabulku obcí zmíněnou v kapitole 5.2. Mimo názvu samosprávních celků a průměrné ceny za metr čtvereční je do výsledné tabulky dopočtena i hodnota množství inzerátů, ze kterých je daný průměr spočítán. U prodeje bytů byly zvlášť provedeny dotazy pro byty ve vlastnictví osobním a družstevním. Po provedení dotazu byla vždy výsledná tabulka exportována z databáze. Názorné dotazy pro pronájem bytů jsou uvedeny níže.

```
SELECT `mesto` , AVG(`cena_za_1m`) , COUNT(id)
FROM `pronajem_byty_cr`
GROUP BY mesto
```

```
SELECT `NAZ_LAU1_P` , AVG(`cena_za_1m`) , COUNT(id)
FROM `pronajem_byty_cr` INNER JOIN Obce ON NAZ_OBEC_P = mesto
GROUP BY NAZ_LAU1_P
```

```
SELECT `NAZ_CNUTS3` , AVG(`cena_za_1m`) , COUNT(id)
FROM `pronajem_byty_cr` INNER JOIN Obce ON NAZ_OBEC_P = mesto
GROUP BY NAZ_CNUTS3
```

Pro seskupení záznamů dle ulic je nutné nejdříve seskupovat dle obcí ve kterých se dané ulice nachází, jelikož stejné názvy ulic v různých obcích jsou poměrně častým jevem. Následné spojení výsledných tabulek s vrstvou reprezentující linie ulic je nutné provést na základě dvou sloupců, a to „*mesto*“ a „*ulice*“. Takový úkon opět použítá GIS aplikace nedovoluje, tyto dva sloupce byly tedy spojeny v rámci databáze ve sloupec „*adresa*“. Obdobné spojení sloupců bylo provedeno i v dbf souboru vrstvy ulic. Následné propojení na základě shody adres je již snadno realizovatelné. Opět je uveden příklad pro pronájem bytů.

```
SELECT AVG(`cena_za_1m`), COUNT(id), CONCAT(mesto,' ',ulice) AS Adresa
FROM pronajem_byty_cr
GROUP BY mesto, ulice
```

Poslední potřebná operace je příprava bodových dat pro interpolaci. K tomu slouží pomocná tabulka bodové vrstvy ulic popsána v kapitole 5.2. Je nutné s pomocí systému řízení báze dat vytvořit takové propojení, aby každý inzerát se zmíněným názvem ulice byl propojen se všemi body dané ulice. Jedná se tedy o vazbu M : N. Zároveň je opět propojení realizováno na základě dvou sloupců, „*ulice*“ a „*mesto*“, tentokrát je však propojení realizováno v rámci databáze a není

potřeba sloupce sjednocovat. Z daných souřadnic X a Y bude následně v GIS aplikaci vygenerována geometrie a vytvořen plnohodnotná bodová vrstva.

```
SELECT X, Y, mesto, ulice, cena, plocha, cena_za_lm, typ_nemovitosti,
budova, vlastnictvi
FROM pronajem_byty_cr INNER JOIN ulice ON mesto = NAZ_OBCE AND ulice =
NAZEV_UL_A
```

7.3. Export dat

Rozhraní phpMyAdmin umožňuje poměrně rozsáhlé možnosti nastavení exportu dat. Export dat z databáze je tedy možné nakonfigurovat tak, aby se jejich formát před zpracováním v prostředí GIS aplikace nemusel již nijak měnit. Vhodným formátem je „CSV pro MS Excel“. Nejdůležitějším parametrem je nastavení správné znakové sady souboru, a to „Windows1250“. V případě užití znakové sady *utf-8*, která by teoreticky také měla být kompatibilní, jsou záznamy s diakritikou z neurčitěho důvodu znehodnoceny. Na závěr je vhodné nechat zapsat na první řádky souborů jména sloupců a verzi Excelu nastavit na hodnotu „Windows“.

Nyní můžeme pozorovat finální výsledky sběru dat. Jedná se o agregované tabulky. Z těch se již dá vyčíst určitý trend v rámci České republiky, či odhalit chyby vzniklé při zpracování. Náhled na ceny nemovitostí dle krajů je zobrazen v tabulce 5. Krajem s nejdražšími nemovitostmi je dle očekávání vždy Hlavní město Praha. Zároveň pokud se jedná o byty, jsou počty inzerátů pro hlavní město výrazně vyšší než tomu je u ostatních krajů. Protipólem Prahy je Ústecký kraj s hodnotami vždy nejnižšími. V hodnotách prodeje bytů se odráží pouze inzerce bytů osobního vlastnictví.

Tabulka 5: Ceny nemovitostí dle krajů

Kraj	prodej byty		pronájem byty		prodej domy	
	Cena za m ² [Kč]	Počet	Cena za m ² [Kč]	Počet	Cena za m ² [Kč]	Počet
Hlavní město Praha	53041	34935	208	78723	45348	4572
Jihomoravský kraj	29484	9759	167	16918	17378	10618
Středočeský kraj	26277	13203	149	8676	24260	26883
Královéhradecký kraj	23618	4128	129	3737	14723	4915
Jihočeský kraj	22162	5752	134	6855	15636	6061
Plzeňský kraj	22039	5243	131	5731	14855	5247
Pardubický kraj	20880	6038	137	4473	13660	5459
Zlínský kraj	20091	3880	143	3718	12527	4517
Olomoucký kraj	19933	4787	130	5363	12419	5554
Karlovarský kraj	19649	4422	128	1497	15361	1980
Liberecký kraj	18909	2734	125	3471	14328	3287
Vysočina	18253	3230	121	1689	13244	4835
Moravskoslezský kraj	16008	7134	122	16324	14882	6809
Ústecký kraj	10897	5970	119	5429	12352	6288

Hodnoty pro okresy reflektují podobný trend jaký lze pozorovat u krajů. Okresy Středočeského a Jihomoravského kraje taktéž dosahují nejvyšších hodnot cen nemovitostí, zatímco okresy Most, Chomutov a Bruntál z Ústeckého a Moravskoslezského kraje jsou nositeli cen nejnižších.

Specifičtější situace nastává při seskupení hodnot na tak malé územní celky jako jsou obce. V tabulce 6 je nastíněna situace prodeje bytů ve vybraných obcích. Rozptyl hodnot je již velký. Za povšimnutí stojí například skutečnost, že Praha je až na sedmé pozici. Na pozicích před ní jsou obce známé tím, že se v nich nachází lyžařské rezorty. Na druhé straně tabulky se pak nachází malé obce s hodnotami vypočtenými pouze z jediného inzerátu nebo jiného velmi malého počtu. Dle fyzické kontroly na serveru Sreality jsou ceny těchto nemovitostí zároveň často podmíněny velmi špatným stavem.

Tabulka 6: Prodej bytů ve vybraných obcích

Obec	Cena za m ² [Kč]	Počet
Špindlerův Mlýn	74588	116
Bedřichov	66210	2
Karlova Studánka	58250	1
Deštné v Orlických horách	55428	2
Pec pod Sněžkou	54412	27
Černý Důl	54393	21
Praha	53041	34935
Čeladná	52778	33
...		
Kelčany	1845	1
Kácov	1687	1
Mokré Lazce	1035	1

Tabulkami nemovitostí seskupenými na jednotlivé ulice, nebo inzeráty s přidanou lokalizací souřadnicemi se momentálně nemá smysl zabývat. Reálné závěry z těchto dat půjde lépe vyvodit až po napojení na prostorovou složku a zobrazení v mapě.

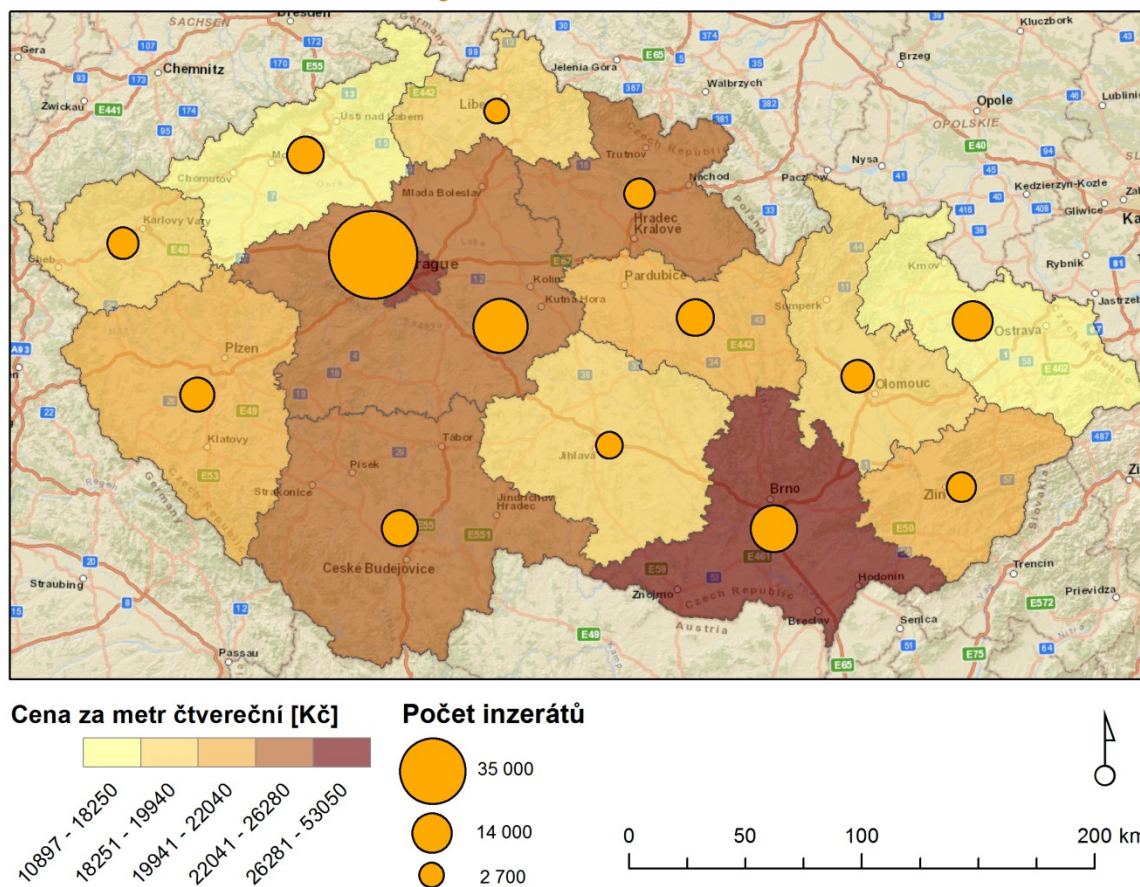
8. Výběr vhodné metody zobrazení, automatizace zpracování

Jak již bylo nastíněno při tvorbě podpůrných tabulek v databázi a při agregaci a výběru dat z databáze, bylo potřeba učinit rozhodnutí, jakým způsobem budou inzertní data ve finální aplikaci prezentována. Dle řešerše stávajících cenových map a dostupných kartografických metod se nabízí hlavně zobrazení cen nemovitostí s využitím kartogramů [14] a ohodnocených linií ve větších měřítcích, nebo pomocí rastrů, jejichž hodnoty jsou vypočteny některou z interpolačních metod. Zároveň je vhodné přípravu dat pro vybrané zobrazení určitým způsobem automatizovat, aby aktualizace zdrojových dat webové aplikace nebyla nepřiměřeně časově náročná. Toto je důležité hlavně pro případné budoucí aktualizace dat.

8.1. Zobrazení pomocí kartogramů a ohodnocených linií

Jako první byla tedy testována možnost užití kartogramů. Data k tomuto účelu již byla exportována z databáze a v prostředí GIS aplikace byla napojena na příslušné vrstvy územních celků. Tyto vrstvy definující území krajů, okresů a obcí byly pro práci poskytnuté Registrem sčítacích obvodů. Stanovení intervalů je vhodné provést užitím kvantilů, dosáhne se tak rovnoměrného zaplnění. Pro přehlednost byla nastavena kartogramům průhlednost a byly podloženy podkladovými mapami uliční sítě ze serveru ArcGIS Online. Ve webové aplikaci bude možnost využít jako podklad mapy Google, disponující vyšší kvalitou a přehledností, než mapy v následujících příkladech. Je samozřejmé, že pro zobrazení v menších měřítcích je vhodnější užít kartogramy na krajích a se zvětšujícím se měřítkem postupně přecházet na menší územní celky. Příklad kartogramu cen prodeje bytů v krajích je uveden na obrázku 11. Pro přehled o počtech inzerátů agregovaných k jednotlivým krajům je v mapě využita metoda kartodiagramu. Vzhledem k rušivosti daného řešení nebudou však kartodiagramy ve finální aplikaci využity.

CENA BYTŮ v krajích v roce 2013

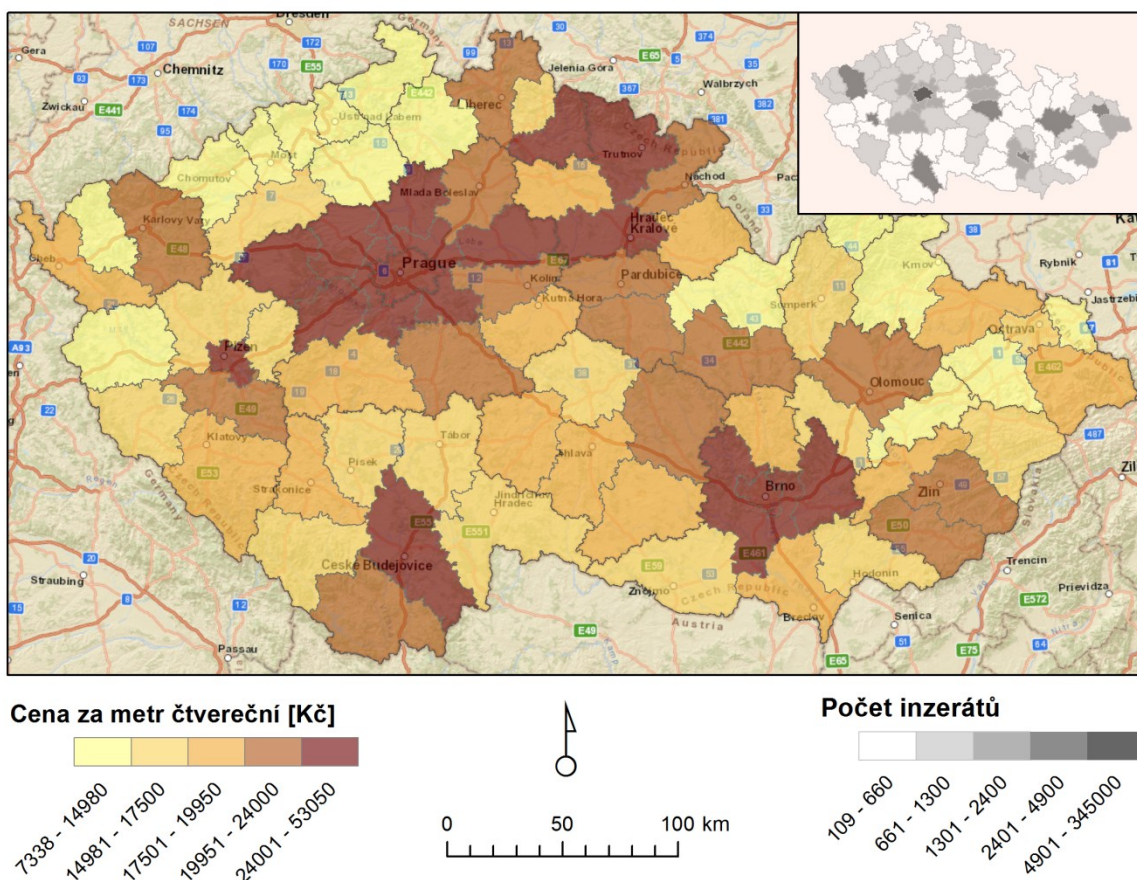


Obrázek 11: Cena bytů v krajích v roce 2013

Z mapy je patrné, že nejdražší byty se nachází v Hlavním městě Praze a v Jihomoravském kraji. Vysoké ceny bytů v Jihomoravském kraji jsou dány hlavně cenami bytů v Brně. Mezi kraje s nejnižšími cenami patří kraj Moravskoslezský a Ústecký. Srovnání počtu inzerčních nabídek, ze kterých je kartogram vytvořen indikuje nejvyšší počet inzerátů v Praze a Středočeském kraji. Nejméně inzerovány jsou byty na Vysočině a v Libereckém kraji.

Po přiblížení a tedy zvětšení měřítka, by webová aplikace měla plynule přejít na území okresů. Stále lze předpokládat existenci inzerce v každém okrese. Lze očekávat větší rozsah hodnot, a proto je nutné definovat nové hodnoty intervalů pro kartogram a legendu. Opět je uveden příklad takového kartogramu pro ceny prodeje bytů na obrázku 12.

CENA BYTŮ v okresech v roce 2013

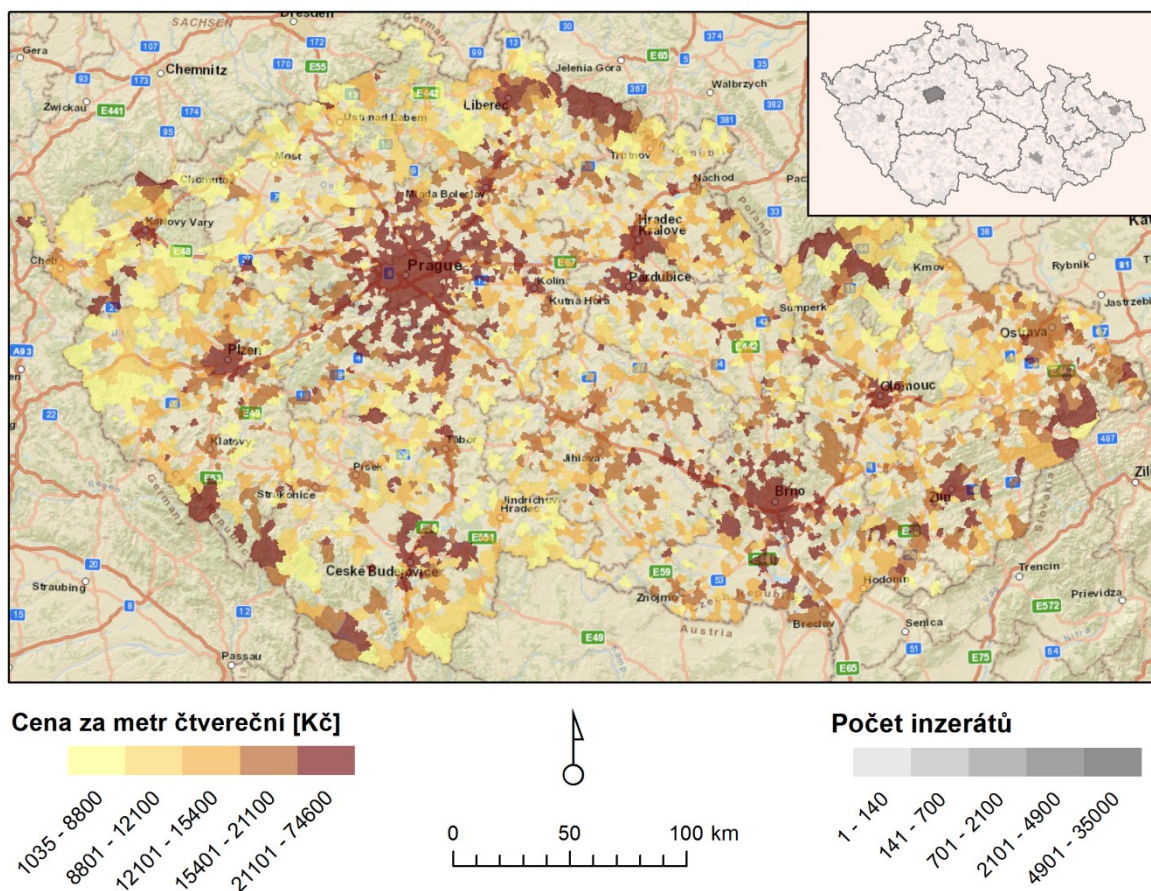


Obrázek 12: Cena bytů v okresech v roce 2013

Vysoké ceny bytů se vyskytují opět v Hlavním městě Praze a v dalších okresech s velkými městy jako Brno, Plzeň, České Budějovice a Hradec Králové, či v blízkosti těchto měst. Výjimkou je pak Ostrava, kde jsou ceny bytů průměrné. Nízké ceny bytů se vyskytují převážně v okresech Moravskoslezského a Ústeckého kraje. Počty inzertních nabídek jsou zobrazeny kartogramem v pravé horní části mapového výstupu. Zcela logicky jsou nejvyšší počty inzerce v okresech velkých měst, s tím že Praha výrazně převyšuje ostatní.

Kartogramem prodeje bytů jednotlivých obcí byla testována možnost zobrazení inzertních dat v druhé nejvyšší měřítkové hladině vyvíjené webové aplikace. Nyní již není území České republiky plně zaplněno daty. V mapě se vyskytují obce bez stanovené hodnoty ceny nemovitostí. To je způsobeno skutečností, že v daných obcích se nevyskytuje daný typ nemovitostí (byty), nebo v ní nebyly nemovitosti inzerovány. Takové obce jsou tedy v kartogramu zastoupeny průhlednou výplní. Příklad seskupení údajů k jednotlivým obcím je uveden na obrázku 13.

CENA BYTŮ v obcích v roce 2013



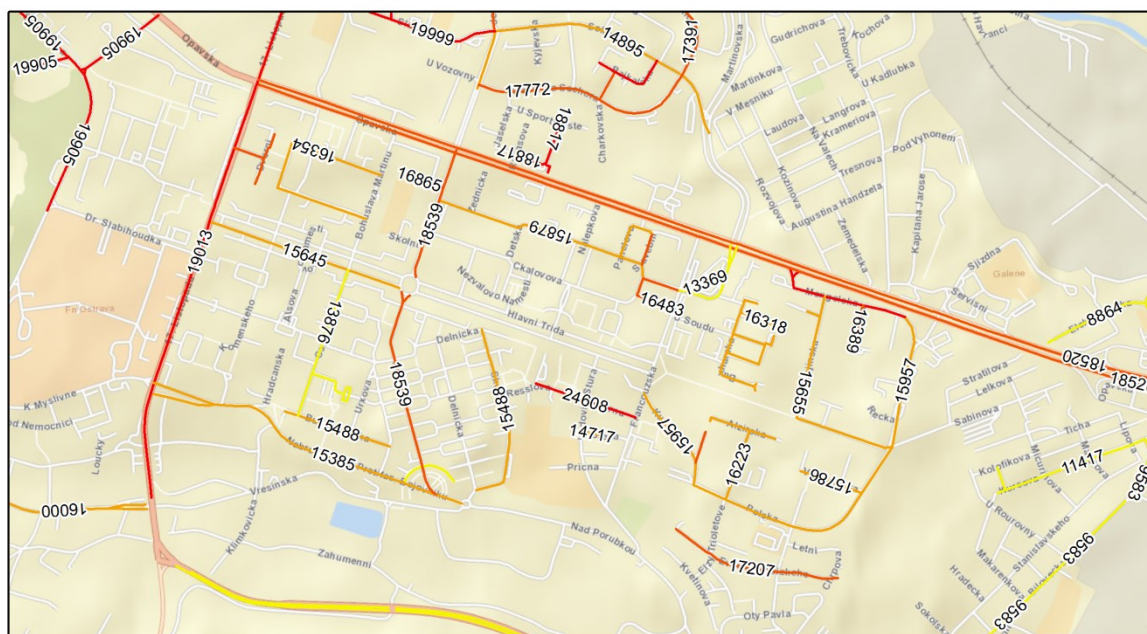
Obrázek 13: Cena bytů v obcích v roce 2013

Jak již bylo zmíněno v kapitole 7.3., ceny bytů při agregování na území jednotlivých obcí již nabývají nejvyšších hodnot v oblastech zvýšeného cestovního ruchu, nejvíce potom v obcích známých jako lyžařské rezorty, například Špindlerův Mlýn, Pec pod Sněžkou, Černý Důl, či Čeladná. Velká města a obce v jejich aglomeracích disponují relativně vysokými cenami taktéž. Nejnížší hodnoty se vyskytují převážně v malých obcích, vzdálených od větších měst, bez občanské vybavenosti.

Pro případ užití plného rozsahu kvality inzertních dat, tedy jejich lokalizace na ulice, je využita metoda zobrazení hodnot dle ulic. K ulicím jsou poté přidány popisky obsahující jejich agregované ceny. Z důvodu zvýšení přehlednosti je vhodné nadefinovat různé barvy linií tak, aby měl uživatel výsledné aplikace hrubý přehled o cenách dříve, než začne číst popisky. Taková situace je zobrazena na obrázku 14. Porovnáním hodnot legendy této mapy s legendou kartogramu krajů lze pozorovat markantní rozdíl v rozsahu hodnot cen nemovitostí. To je způsobeno tím, že při agregaci na menší geoprvky (v tomto případě ulice) vstupují

extrémně nízké či vysoké hodnoty některých inzertních nabídek do výpočtu průměru s žádným nebo malým počtem ostatních nabídek.

CENY BYTŮ dle ulic v roce 2013 v Ostravě Porubě



Cena za metr čtvereční [Kč]

- 8864 - 14700
- 14701 - 16360
- 16361 - 18540
- 18541 - 24700

0 0,5 1 km



Obrázek 14: Ceny bytů dle ulic v roce 2013 v Ostravě Porubě

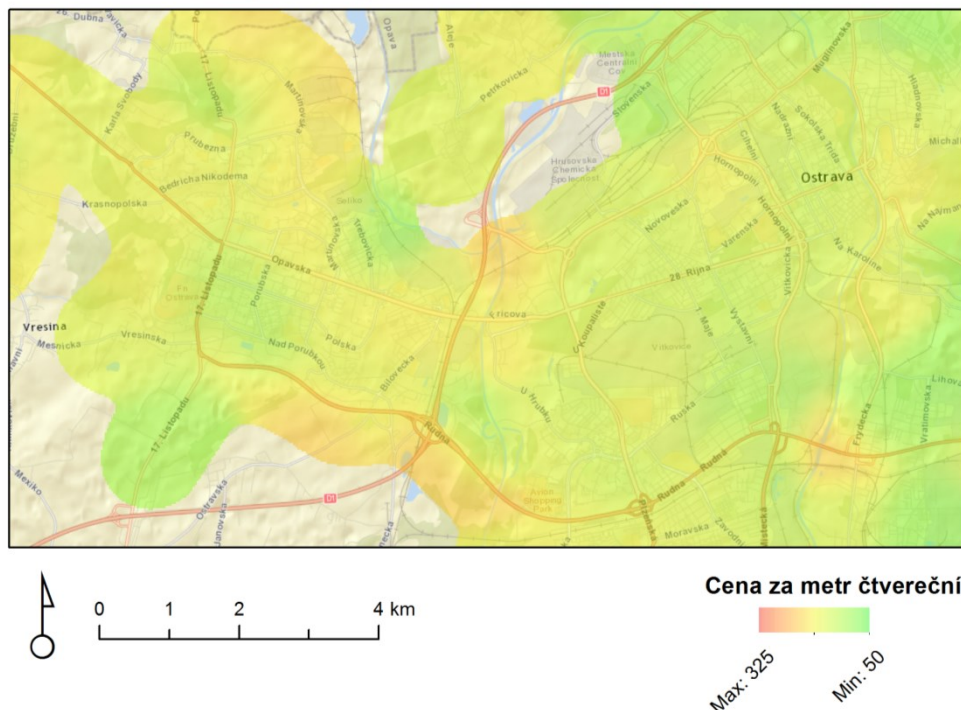
8.2. Zobrazení pomocí rastrů

Druhou možností, hojně využívanou u ostatních cenových map je zobrazení pomocí rastrů z interpolovaných hodnot. Organizace, od kterých tyto cenové mapy pochází, však disponují přesně lokalizovanými hodnotami transakčních cen, zatímco zdroj dat pro tuto práci umožňuje lokalizaci nejlépe na název ulice a obce. Zároveň by data vstupující do interpolačního výpočtu měla být bodového charakteru.

Bylo tedy nutné vyvinout způsob jak tyto nedostatky zpracovávaného zdroje dat obejít. Z toho důvodu byla v databázi zřízená dříve popsaná tabulka ulic, reprezentovaných body a data byly z databáze vhodně exportována. V užití GIS aplikaci musela být z těchto dat nejdříve vygenerována bodová vrstva. Teprve z té lze interpolovat. Vzhledem k výpočetní náročnosti tohoto úkolu, byly jako vstup do interpolace zvoleny inzertní body pouze za území Ostravy. Využita byla difuzní

interpolace s nastavením velikosti buňky výstupního rastru 20 metrů a maximální vzdáleností, do které jsou vyhledávány okolní body, 300 metrů. Daný typ interpolační metody byl využit za účelem vyhlazení extrémů a skokových rozdílů mezi jednotlivými inzertními nabídkami. Výstupnímu rastru byla taktéž nastavena průhlednost a byl podložen podkladovou mapou. Výsledek pro pronájem bytů je zobrazen na obrázku 15.

Pronájem bytů v Ostravě za rok 2013

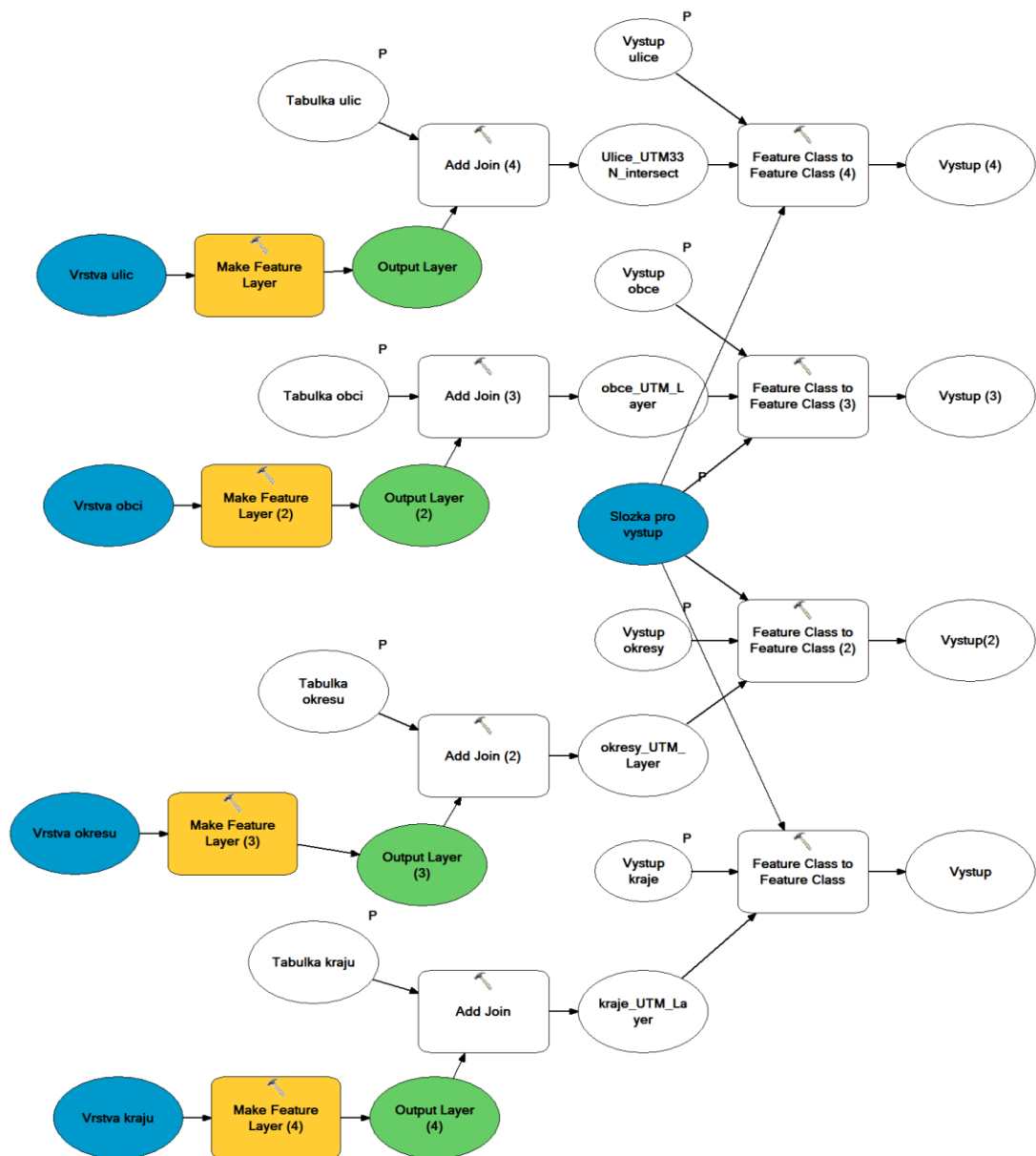


Obrázek 15: Zobrazení pomocí interpolace

Ačkoliv může být výsledek považován za přijatelný, ze širšího hlediska jej využít nelze. Nejvíce limitující je právě skutečnost, že bodové údaje pro interpolaci jsou zde jakýmsi způsobem pouze simulovány z uliční sítě. Pokud se na ulici nachází více inzerátů, jsou umístěny vždy do stejných bodů reprezentující danou ulici, místo toho aby reprezentovaly přesnou lokaci dané nemovitosti. Z takového nakupení bodů různých hodnot na stejných místech často vznikají při interpolaci bodové fenomény extrémů. V rámci jednoho města lze sice parametry interpolace nastavit tak, aby byl výsledek vizuálně přijatelný, pokud však chceme interpolovat území celé republiky, zjistíme, že nelze nalézt kompromis, se kterým by výstupní rastr byl přijatelný jak pro hustě zalidněné města s rozsáhlou inzercí nemovitostí, tak po malé obce s několika inzeráty. Taktéž nesmíme zapomínat na výrazně výpočetně náročnější výpočet interpolace. Z těchto důvodů bylo od metody zobrazení cen nemovitostí pomocí rastrů ustoupeno a byla zvolena pouze metoda zobrazení pomocí kartogramů a ohodnocených linií.

8.3. Automatizace propojení dat

Propojení datových tabulek exportovaných z databáze s vektorovými shapefile soubory není sice složitý proces, v případě následných aktualizací vyvíjené aplikace by však časová náročnost jeho ručního zpracování byla limitující. Z toho důvodu byl v prostředí ESRI ArcMap vyvinut pomocí nástroje ModelBuilder toolbox s požadovanou funkčností propojení dat. Struktura modelu je zobrazena na obrázku 16. Po spuštění nástroje je uživatel pouze dotázán na umístění tabulek agregací krajů, okresů, obcí a ulic a na názvy výstupních souborů. Následné propojení a export do výstupní složky je provedeno automatizovaně a ušetří uživateli čas, který by strávil ručním zpracováním.

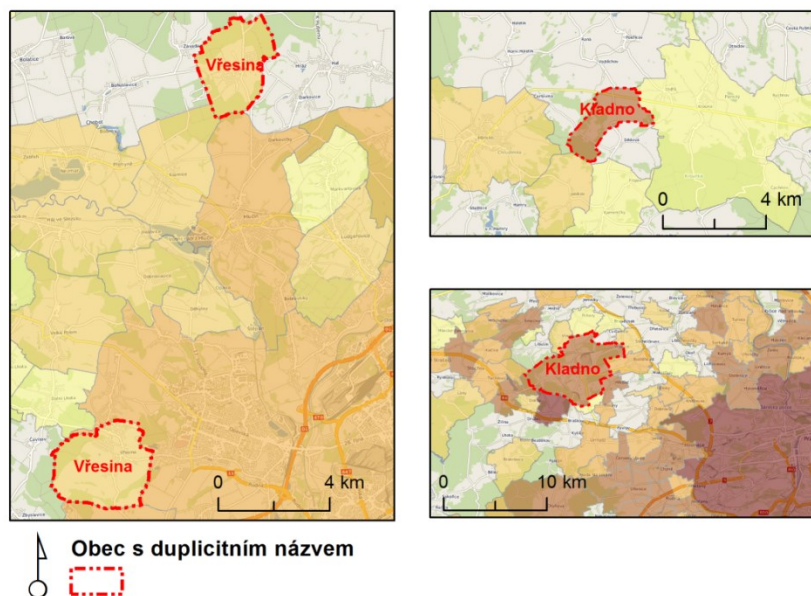


Obrázek 16: Model zpracování propojení dat

8.4. Nedostatky plynoucí ze zdroje dat

Jak již bylo mnohokrát zmíněno, inzeráty jsou prostřednictvím RSS kanálu lokalizovány pouze na název obce a ulice. Jakákoliv unikátnější informace typu kód obce či poštovní směrovací číslo chybí. Zároveň v rámci RSS kanálu není uvedena příslušnost obce k vyššímu územnímu celku. Z toho důvodu vznikají dvě poměrně závažné chyby, které bohužel nelze momentálně vyřešit. První chybou je duplicitní přidělování inzerátů obcím České republiky, které nesou stejný název. Nelze totiž automatizovaně stanovit, které z obcí dané inzeráty připadají. Příklady tohoto nedostatku jsou uvedeny na obrázku 17. Nejčastěji se jedná o malé obce shodných názvů (Nová Ves, Lhotka atd.), horší situace nastává ale u názvů jako Kladno či Jesenice. Ty totiž reprezentují poměrně zalidněné obce, ale zároveň i malé vesnice. Tyto vesnice získají propojením dat údaje patrně patřící jejich větším jmenovcům. Jednou z možností, jak tento problém řešit je odstranění všech menších obcí s duplicitním názvem z databáze. Daleko lepším řešením by však bylo po dokončení práce a předvedení výsledků pověřeným osobám z portálu Sreality, apelovat na dané osoby s prosbou o rozšíření RSS kanálu například o okres či ORP, které by umožnily přesnou identifikaci obcí.

Příklady obcí s duplicitními názvy



Obrázek 17: Obce s duplicitními názvy

Druhou méně závažnou chybou jsou dlouhé ulice. Při agregaci takové ulice nesou průměrnou cenu ovlivněnou údaji například z jiné městské části dané obce. Problém je výraznější hlavně v oblastech, kde při dané ulici není žádná zástavba. Částečným řešením by v budoucnu mohlo být ořezání vrstvy ulic určitou polygonovou vrstvou zastavěného území. V době tvorby diplomové práce taková datová vrstva však dostupná nebyla.

9. Tvorba webových mapových aplikací

Po vytvoření všech potřebných mapových vrstev zbývalo vyvinout finální webovou aplikaci, ve které budou vrstvy zpřístupněny. Nabízely se možnosti užití proprietárního software ESRI ArcGIS Server nebo Open Source řešení formou aplikace Geoserver a knihoven OpenLayers. Vyvíjená aplikace bude určena pouze pro prohlížení dat a nebude obsahovat složité výpočetní geoinformatické funkce lépe dostupné právě v knihovnách software ESRI, a tedy není důvod přistupovat k užití proprietárního softwaru. Výhodou Open Source řešení je pak skutečnost, že kombinace Geoserveru a OpenLayers je výrazně méně výpočetně náročná a funguje i pod snadněji přístupným systémem Linux.

Aplikace Geoserver je ve vybraném řešení využita pro uložení mapových vrstev a jejich publikaci prostřednictvím WMS (webové mapové služby), zatímco knihovny OpenLayers jsou užity k zobrazování map přímo ve webovém prohlížeči.

Jelikož je potřebné, aby výsledná webová aplikace byla uživatelům trvale dostupná v prostředí webu, je nutné umístit webové stránky aplikace, knihovny OpenLayers, ale i samotný Geoserver na patřičný server. Dříve užitá řešení studgisweb, používané pro automatizované stahování dat do prostředí databáze, již není dostačující. Z toho důvodu byl Insitutem Geoinformatiky zřízen virtuální server určený pro tuto práci. Server opět disponuje sadou programového vybavení LAMP popisovanou v kapitole 5. Zároveň však již obsahuje nainstalovanou aplikaci Geoserver a uživatel má nastavené právo „sudo“ (substitute user do). Sudo je příkaz užívaný v unixových systémech. Slouží k vykonání operace s oprávněními jiného uživatele, jímž je v tomto případě root. Bez tohoto práva by nebylo možné uložit na server mapové vrstvy, upravit nastavení proxy nebo provádět jiné úkony nutné k dokončení webové aplikace.

9.1. Publikace vrstev prostřednictvím aplikace Geoserver

Jak již bylo zmíněno, Geoserver je Open Source aplikace, napsána v programovacím jazyce Java. Umožňuje sdílet, ale i upravovat geografická data. Klade důraz na interoperabilitu, data publikuje s využitím otevřených standardů v oblasti. Kromě v této práci využívaného standardu WMS, dokáže taktéž implementovat standardy WCS (web coverage service) či WFS (web feature service). [5] Většina pracovních úkonů lze provést pomocí webového rozhraní, které aplikace poskytuje.

Prvním krokem k publikaci mapových vrstev je vytvoření složky v adresáři data_dir. Do té musí být následně nahrány všechny vrstvy cen nemovitostí v jednotlivých agregacích. V tomto případě budou publikovány vrstvy cen pronájmu bytů, prodeje bytů osobního vlastnictví a prodeje domů. Ke každému

typu nemovitosti připadají čtyři agregace – za kraje, okresy, obce a ulice. Celkem tedy bylo importováno 12 vrstev.

Následné úkony lze již provádět ve webovém rozhraní Geoserveru. Nejdříve je potřeba vytvořit určitý pracovní prostor. Tento pracovní prostor je určen k seskupení vrstev shodného zájmu, v tomto případě cenových map. Veškeré nastavení či styly daných vrstev pak budou uloženy právě v daném pracovním prostoru a v případě migrace mapové aplikace bude její přesun výrazně usnadněn.

Následuje vytvoření datového skladu. Zdrojem dat daného skladu může být soubor, skupina souborů, databázová tabulka, rastrový soubor, nebo přímo složka souborů. Smyslem datového skladu je definování parametru připojení pouze jednou pro celou skupinu dat v něm, nikoli pro jeho každou část zvlášť. V tomto případě se jedná o datový sklad vektorových souborů, přesněji pak „Složku prostorových souborů (shapefiles)“. Při vytváření je nutné nastavit pracovní prostor, české kódování znaků Windows1250 a jako zdroj dat dříve vytvořenou složku s nahranými datovými vrstvami.

Po vytvoření datového skladu je možné přistoupit k samotnému publikování vrstev. Publikování uživatel aktivuje přidáním nového zdroje v záložce vrstev. Po výběru správného datového skladu a pracovního prostoru jsou zobrazeny všechny soubory v daném skladu. Každý soubor je pak potřeba publikovat zvlášť. Při publikaci je nutné zvolit prostorový referenční systém. V tomto případě „ETRS UTM Zone 33N“. Další nedílnou součástí publikace je definování ohraničujících rámců publikovaných vrstev, ty lze vygenerovat ze zdrojových dat. Vyplnění metadat je rovněž vhodné.

V tento moment jsou vrstvy vypublikovány se standardní symbologií. Pro dosažení vytyčeného vzhledu vrstev z kapitoly 8.1. je nutné vytvořit styly zobrazení. V prostředí aplikace Geoserver je stylování dosaženo prostřednictvím značkovacího jazyka „Styled Layer Descriptor“ (SLD styly). SLD je značkovací jazyk založený na XML, umožňující poměrně rozsáhlé definice symbologie, i když poněkud složité.

Příklad syntaxe jazyka SLD je uveden v následujícím fragmentu kódu. Jedná se pouze o definování jednoho intervalu hodnot kartogramu prodeje bytů v krajích. Hlavním prvkem je uvedení hranic intervalu a definice symbologie polygonů, tedy jejich ohraničení a výplně. Dále je definována měřítková hladina, při které se vrstva zobrazí a název či titulek vrstvy, ty jsou použité při automatizovaném generování legendy metodou `getLegendGraphic`. Hranice intervalů jsou vypočteny dle kvantilů a zaokrouhleny na jednotky při pronájmu nemovitostí a stovky při jejich prodeji.

```
<Rule>
  <MinScaleDenominator>1000001</MinScaleDenominator>
  <Name><![CDATA[18301 - 20000]]></Name>
  <Title><![CDATA[18301 - 20000]]></Title>
  <ogc:Filter>
    <ogc:PropertyIsBetween>
      <ogc:PropertyName>prum_cena</ogc:PropertyName>
      <ogc:LowerBoundary>
        <ogc:Literal>18301</ogc:Literal>
      </ogc:LowerBoundary>
      <ogc:UpperBoundary>
        <ogc:Literal>20000</ogc:Literal>
      </ogc:UpperBoundary>
    </ogc:PropertyIsBetween>
  </ogc:Filter>
  <PolygonSymbolizer>
    <Fill>
      <CssParameter name="fill" >#fad155</CssParameter>
      <CssParameter name="fill-opacity" >0.65</CssParameter>
    </Fill>
    <Stroke>
      <CssParameter name="stroke" >#6e6e6e</CssParameter>
      <CssParameter name="stroke-width" >0.4</CssParameter>
      <CssParameter name="stroke-opacity" >0.65</CssParameter>
    </Stroke>
  </PolygonSymbolizer>
</Rule>
```

Celý kód SLD stylu se skládá z pěti podobných pravidel definujících dané intervaly kartogramu a hlavičkových údajů na začátku souboru. V případě ohodnocených linií je pak kód ještě rozšířen o definici popisků daných linií, tedy jejich obsah i formátování. Takový styl musí být vytvořen zvlášť pro každou mapovou vrstvu. Celkem bylo potřeba napsat dvanáct SLD stylů, uložit je v prostředí aplikace Geoserver a přiřadit je k odpovídajícím vrstvám. Všechny styly jsou uloženy na přiloženém CD v rámci pracovního prostoru vyvíjené aplikace.

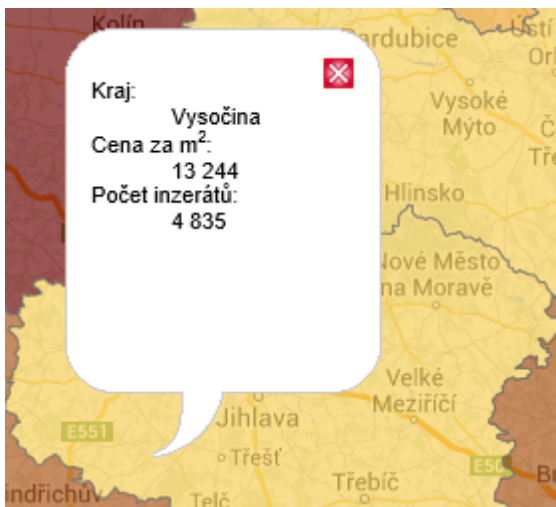
Aby bylo dosaženo žádaného pyramidového efektu zobrazení vrstev, kdy v nejmenším měřítku jsou zobrazeny kartogramy krajů a při zvětšování měřítko se vrstvy postupně mění, až na samotné linie ulic, je vhodné seskupit vrstvy do skupin. Ve skupině vrstev lze definovat pozici vrstev z hlediska překryvu, tento problém však již řeší nastavení měřítkových hladin v SLD stylech. Stejně jako u jednotlivých vrstev je zde nutné vypočítat nebo nastavit ohraničující rám dat a v případě potřeby odlišný souřadnicový systém.

Poslední operací je definice vzhledu odpovědi na dotaz GetFeatureInfo zaslaných uživatelem na server. Tyto dotazy jsou užívány ke zjištění hodnot atributů geografických objektů zobrazených ve výsledné mapové aplikaci. Vzhled odpovědi je definován prostřednictvím systému šablon Freemarker. Tento systém je rovněž využíván k přizpůsobení vzhledu GeoRSS a KML. Opět je potřeba ke každé publikované vrstvě přiřadit šablonu. Tentokrát již nelze pracovat

prostřednictvím webového rozhraní. Vzhled je definován souborem content.ftl, který je potřeba vytvořit ve složce dané vrstvy v pracovním prostoru. Příklad definice dané šablony pro kartogram prodeje domů v krajích je uveden níže. Všechny šablony jsou v rámci pracovního prostoru opět uloženy na příloženém CD.

```
<#ftl encoding="utf-8">
<#list features as feature>
<dl class="dl-horizontal">
<dt>Kraj:</dt>
<dd class="text-info">${feature.NAZ_CNUTS3.value}</dd>
<dt>Cena za m<sup>2</sup>:</dt>
<dd class="text-info">${feature.prum_cena.value?number?round}</dd>
<dt>Počet inzerátů:</dt>
<dd class="text-info">${feature.pocet.value?number?round}</dd>
</dl>
</#list>
```

Šablona je definována jako seznam podmínek a popisů. V podmínkách jsou uvedeny statické názvy jako název typu územního celku, „cena za m²“ a „počet inzerátů“, zatímco v popisech jsou uváděny odpovídající dynamické údaje. Výsledek následné odpovědi na atributový dotaz lze spatřit na obrázku 18.



Obrázek 18: Výsledek atributového dotazu

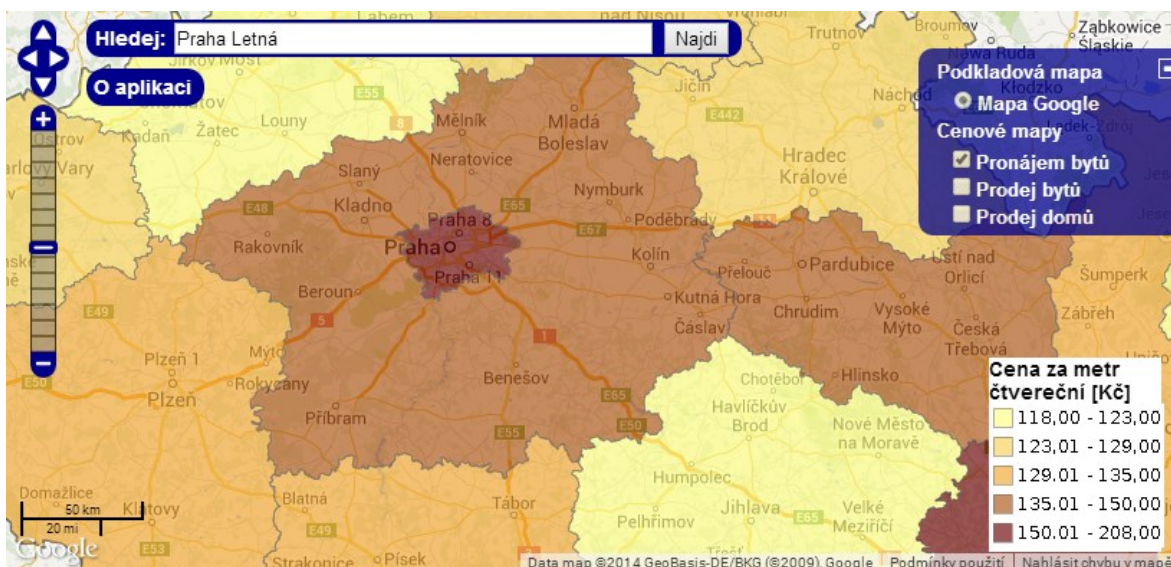
9.2. Vývoj aplikace s užitím knihoven OpenLayers

Jak již bylo zmíněno, OpenLayers jsou otevřená javascriptová knihovna určená k zobrazování map ve webovém prohlížeči. Poskytuje API pro tvorbu webových mapových aplikací. Použita je například pro zobrazování map projektu OpenStreetMap a je základem pro vývoj webové mapové aplikace k této diplomové práci.

Pro plnou funkcionalitu výsledné aplikace byl užitý mimo samotné knihovny OpenLayers jazyk html, CSS, javascript a jQuery. Jazyk html je hlavním z jazyků pro vytváření stránek v systému world wide web, který umožňuje publikaci dokumentů na internetu. CSS neboli kaskádové styly jsou určeny pro popis způsobu zobrazení stránek napsaných v jazycích html, xhtml či xml. Javascript je multiplatformní objektově orientovaný skriptovací jazyk, který se nyní nejčastěji používá jako interpretovaný programovací jazyk pro www stránky a jsou jím zpravidla ovládány různé interaktivní prvky grafického rozhraní. Poslední jQuery je javascriptová knihovna, která klade důraz na interakci mezi javascriptem a html.

Vyvíjená webová mapová aplikace je tedy zjednodušeně řečeno webová stránka napsána v jazyce html. Mapové vrstvy jsou k ní připojeny prostřednictvím knihoven OpenLayers z aplikace Geoserver. Nadstandardní ovládací prvky či přizpůsobení aplikace jsou naprogramovány v jazyce javascript a jQuery. Kaskádové styly jsou pak využity k definování vzhledu a umístění nadstandardních prvků. Vzhledem k rozsáhlosti řešení a decentralizaci funkcí v této kapitole nebudou vypisovány úryvky kódu. Celý html dokument s plnou funkcionalitou je však uložen na přiloženém CD k této diplomové práci.

V hlavičce html dokumentu se nachází kromě standardního nastavení kódování, názvu stránky a autora import využitých knihoven, nadefinované kaskádové styly pro jednotlivé prvky stránky a množství kódu v jazyce javascript a jQuery uskutečňující chod aplikace. Samotné tělo dokumentu obsahuje pouze výpis neutrálních značek <div>. Ty uživateli poskytují mapové okno se standardními ovládacími prvky OpenLayers, legendu, vyhledávání adres a odkaz na okno s informacemi o aplikaci. Jejich funkčnost je definována kódem z hlavičky, či importovanými knihovnami. Náhled na výslednou mapovou aplikaci je poskytnut na obrázku 19.



Obrázek 19: Výsledná mapová aplikace

Standardními mapovými prvky, které jsou poskytnuty knihovnou OpenLayers je samotné mapové okno, nástroj pro přiblížení a posun mapy, měřítko a okno s výběrem mapových vrstev. Formulář pro vyhledávání adres, interaktivní legenda a tlačítko pro zobrazení základních informací o aplikaci musely být doprogramovány.

Jak již bylo zmiňováno, připojení skupinových vrstev pronájmu bytů a prodeje bytů a domů je realizováno prostřednictvím webové mapové služby (WMS) aplikace Geoserver. Jako podkladová mapa je využito aplikační rozhraní Google map. Výběr mapových vrstev byl upraven tak, aby nebylo možné zobrazit více cenových map v jeden moment.

Kritickým úkolem bylo zobrazení dynamické legendy. Požadavkem bylo, aby byla legenda zobrazena zvlášť ke každé mapové vrstvě, ale také zvlášť ke každé aktivní agregaci. To znamená, pokud je například právě aktivní vrstva kartogramu prodeje bytů v krajích, aby byla zobrazena legenda pouze pro tento kartogram, po přiblížení na kartogram prodeje bytů v okresech byla změněna a po přepnutí vrstvy došlo ke změně taktéž. Příklad dynamiky při změně měřítka je uveden na obrázku 20.



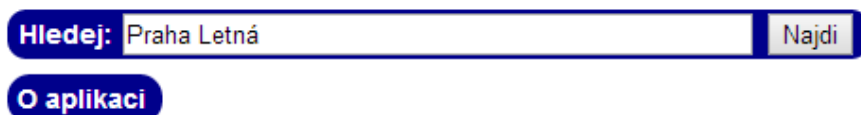
Obrázek 20: Příklad změny hodnot legendy pronájmu bytů při změně měřítka

Samotnou legendu lze vygenerovat pomocí dotazu GetLegendGraphics zaslaného aplikaci Geoserver. Ta následně vytvoří legendu podle SLD stylu, jehož tvorba byla popsána v kapitole 9.1. a zašle zpět obrázek legendy. Knihovny OpenLayers však nedokážou indikovat, která z dílčích vrstev dané skupiny nemovitostí je v dané měřítkové hladině právě zobrazována, a tedy není možné určit, o kterou legendu je potřeba požádat. Tento problém byl řešen zjištěním aktuální hladiny přiblížení, dle které byly následně generovány požadavky na zaslání legendy. Funkci pro zaslání požadavku o legendu pak spouští události přiblížení a změny skupinové vrstvy.

Aby koncový uživatel nebyl omezen při určení hodnoty ceny nemovitostí pouze legendou, či popiskem u linií ulic, byla v aplikaci dodána funkce zjištění dané ceny kliknutím na vybraný územní celek, či ulici. Mimo samotnou cenu uživatel zjistí i počet inzerátů, ze kterého byl daný průměr vypočten. Dotaz je

proveden žádostí GetFeatureInfo na server a jeho formátovaná odpověď je popsána na konci kapitoly 9.1.

Na závěr byl k aplikaci dodán formulář pro vyhledávání adres, který po zadání adresy a zaslání požadavku na vyhledání provede přiblížení na danou lokalitu. Geokódování pro danou funkcionalitu poskytuje portál openrouteservice.org. Implementováno bylo rovněž tlačítko pro zobrazení informací o dané aplikaci. Náhled těchto prvků lze spatřit na obrázku 21.



Obrázek 21: Vyhledávání adres a informace o aplikaci

9.3. Porovnání aplikace s konkurencí

V závěru vývoje by měla být aplikace srovnána se svou konkurencí, aby bylo možné vývoj uzavřít a přejít k celkovému zhodnocení. Původní myšlenka byla aplikaci srovnat se všemi volně dostupnými cenovými mapami na českém trhu. V období psaní této kapitoly však byla aplikace Státního fondu rozvoje bydlení již delší dobu nedostupná a cenová mapa ze serveru Bezrealitky taktéž. Srovnání mohlo být tedy provedeno pouze s aplikací Asociace realitních kancelářů.

Asociace realitních kancelářů disponuje přesně lokalizovanou inzercí nemovitostí a skutečnými cenami prodeje daných nemovitostí, z toho hlediska je její výchozí pozice mnohem lepší, než u této diplomové práce. Díky svému zdroji dat, tedy určité databázi inzertních nabídek byla schopna vybudovat cenové mapy formou interpolovaných rastrů, zatímco tato práce se musela spokojit s metodou kartogramů a ohodnocených ulic. Celkovým vzhledem a funkcionalitou tuto diplomovou práci však nepřevyšuje. Co se týče přehlednosti zobrazovaných dat, metody vykreslení rastrů a zvolené legendy, je možné říci, že je dokonce horší. Velkou výhodou aplikace vyvinuté při této diplomové práci je možnost zjištění z kolika inzerátů byla zobrazovaná cena v mapě vypočtena. Tuto informaci asociace neuvádí a neuváděl jí ani server bezrealitky.cz.

Mimo záležitosti vzhledu byly porovnávány i samotné ceny. Několik cen prodeje bytů v namátkou zvolených lokalitách je uvedeno v tabulce 7. Celkově si hodnoty cen prodeje bytů odpovídají, nedá se tedy říci, že by existovaly vážné rozdíly. V mnoha případech však ceny Asociace realitních kancelářů, jak lze spatřit i v dané tabulce, určitým způsobem převyšují hodnoty zjištěné z naší aplikace. Vzhledem ke skutečnosti, že ceny využití touto prací jsou ceny inzertní, a tedy patrně vyšší než ceny reálně uskutečněné, je tento jev zajímavý. Mohlo by to indikovat umělé navyšování cen realit cenovou mapou, aby stávající nabídky

vypadaly lukrativněji. Druhým důvodem by mohl být malý počet inzertních nabídek s vyšší cenou.

Tabulka 7: Porovnání cen bytů ve vybraných lokalitách

Lokalita	Vyvíjená aplikace [Kč/m ²]	Asociace realitních kanceláří [Kč/m ²]
Praha - Veletržní	52509	59399
Ostrava - 17. listopadu	19113	18933
Havířov - Dlouhá třída	13933	17094
Karviná - Borovského	10391	12513

Zajímavější je však srovnání cen prodeje domů. Zde Asociace realitních kanceláří viditelně zaostává. Opět je uvedeno několik příkladů v tabulce 8. U lokalit hustší domovní zástavby cenová mapa vykazuje podobné výsledky jako při prodeji bytů. Z rastru její cenové mapy však ve velkém měřítku lze vyčíst, že je vypočten jen z malého množství hodnot. Velké množství menších obcí zde potom nedisponuje žádnou vypočtenou hodnotou. Extrémně nízká hodnota u obce Horní Suchá je patrně způsobena skutečností, že se jedná o jedinou hodnotu v dané lokalitě, navíc patrně nemovitosti nízké kvality.

Tabulka 8: Porovnání cen domů ve vybraných lokalitách

Lokalita	Vyvíjená aplikace [Kč/m ²]	Asociace realitních kanceláří [Kč/m ²]
Ostrava Třebovice	18075	bez inzerce
Praha Dejvice - Šárecká	74393	75578
Albrechtice u Českého Těšína	12938	bez inzerce
Horní Suchá	13775	2024

Shrneme-li poznatky z celkového srovnání aplikace vyvinuté při této diplomové práci a komerční aplikace Asociace realitních kanceláří, dostáváme závěr, že cenová mapa asociace se sice může jevit estetičtěji, nicméně po praktické stránce vyvíjení cenovou mapu nepřevyšuje a v aspektech zaplnění daty dokonce zaostává.

Tabulka 9: Shrnutí cenové mapy vyvinuté při této diplomové práci

Klady	Zápory
Mapové vrstvy sestaveny z velkého množství inzertních nabídek	Lokalizace nejpřesněji na název ulice a obce
Plně popsána metodika zpracování zdrojových dat a časový rámec	Rychlost odezvy a vykreslování vrstev v aplikaci je z důvodu slabšího serveru pomalejší
Možnost zjištění počtu inzerátů, ze kterých byla spočtena průměrná hodnota uvedená v mapě	Zdroj dat zatížený chybami, které je nutné filtrovat

10. Závěr

Hlavní cíl této diplomové práce, tedy vytvoření webové mapové aplikace cenové mapy pro prodej a pronájem nemovitostí v České republice byl splněn. Cenová mapa je dostupná na adrese gis.vsb.cz/nemovitosti.

Údaje o cenách nemovitostí byly získávány prostřednictvím RSS kanálu realitního portálu Sreality. Navzdory určitým nedostatkům tohoto kanálu pro potřeby zpracovávané diplomové práce, byla vytvořena aplikace realizující stahování inzertních nabídek a jejich import do prostředí databáze. V té došlo k filtraci chybných údajů a následnému exportu k finálnímu zpracování pro zobrazení dle vybrané kartografické metody. Výsledné mapové vrstvy byly poté využity k prezentaci cen nemovitostí ve vyvinuté webové mapové aplikaci.

Ačkoliv byla výchozí pozice z pohledu zdroje dat pro vyvíjenou aplikaci poměrně nepříznivá, povedlo se právě vhodným zpracováním RSS kanálu, filtrací dat a výběrem kartografické metody při tvorbě výsledných mapových vrstev dosáhnout uspokojivých výsledků. Jediným významným nedostatkem zůstává nejednoznačné přidělení inzertních údajů obcím se stejným názvem. Tento nedostatek je způsoben samotným zrojem dat a jeho odstranění nebylo možné. V budoucnu, po předvedení výsledků diplomové práce zástupcům společnosti Seznam, by teoreticky mohl být zpracováváný RSS kanál obohacen o údaje, které by tento nedostatek odstranily.

Z hlediska srovnání s konkurenčními aplikacemi je cenová mapa vyvinutá touto diplomovou prací taktéž úspěšná. Dvě ze tří aktivních českých aplikací během tvorby této diplomové práce přestaly být funkční a třetí, tedy cenová mapa Asociace realitních kanceláří vykazuje srovnatelné, ne-li horší kvality. Jako jediná aplikace vyvinutá při této diplomové práci nabízí možnost zjištění z kolika nabídek je průměrná cena nemovitostí v dané oblasti vypočtena.

Pro případné laické uživatele lze tedy cenovou mapu doporučit, minimálně do té doby, než se na české scéně objeví kvalitnější aplikace vycházející z přesnějšího a důvěryhodnějšího zdroje dat, než jsou inzertní nabídky realitních kanceláří. Delší užívání této cenové mapy je však podmíněno alespoň každoroční aktualizací zobrazovaných dat, pro kterou jsou připraveny postupy takovou aktualizaci urychlující.

Seznam zdrojů

- [1] Sreality.cz. [online]. [cit. 2013-10-08]. Dostupné na WWW: <<http://www.napoveda.seznam.cz/cz/sreality/>>.
- [2] HORÁK, J.: *Nové internetové datové zdroje*, Institut geoinformatiky, HGF VŠB-TU Ostrava, 31 s.
- [3] CUPAL, M.: *Porovnávací (komparativní) metody oceňování nemovitostí*, Ústav soudního inženýrství VUT v Brně
- [4] PEREZ, A. S.: *OpenLayers Cookbook*, Packt Publishing 2012, 300 s. ISBN: 1849516685.
- [5] IACOVELLA, S.: *GeoServer Beginner's Guide*, Packt Publishing 2013, 350 s. : ISBN: 1849517843.
- [6] *Cenová mapa bezrealitky.cz*. [online]. [cit. 2014-09-03]. Dostupné na WWW: <<http://cenovamapy.bezrealitky.cz/>>
- [7] *Cenová mapa Asociace realitních kancelářů*. [online]. [cit. 2014-09-03]. Dostupné na WWW: <<http://cenovamapa.gekonsro.cz/>>
- [8] *Cenová mapa Státního fondu rozvoje bydlení*. [online]. [cit. 2014-09-03]. Dostupné na WWW: <<http://www.sfrb.cz/servis/mapa-najemneho.html>>
- [9] *National home prices page*. [online]. [cit. 2014-09-03]. Dostupné na WWW: <http://www.trulia.com/home_prices/>
- [10] *Cenová mapa transakčních cen*. [online]. [cit. 2014-26-03]. Dostupné na WWW: <<https://www.cenovamapa.org/>>
- [11] *Definice vybraných ukazatelů bytové výstavby*. [online]. [cit. 2013-10-08]. Dostupné na WWW: <http://www.czso.cz/xb/redakce.nsf/i/metodicke_vysvetlivky_definice_vybranych_ukazatelu_bytove_vystavby>
- [12] VONDRÁKOVÁ, A.: *Autorské právo v kartografii a geoinformatice*, Katedra geoinformatiky UP Univerzita Palackého v Olomouci, 126s. ISBN: 978-80-244-3205-2
- [13] SCOUARNEC, Y. L.; BORONCZYK; KOLEKTIV: T.: *PHP 6, MySQL, Apache*, Computer Press 2009, 816 s. ISBN: 978-80-251-2767-4.
- [14] VOŽENÍLEK, V.; KAŇOK, J.; KOLEKTIV: *Metody tematické kartografie* Univerzita Palackého v Olomouci, 216 s. ISBN: 9788024427904.

Seznam obrázků

Obrázek 1: Cenová mapa bezrealitky.cz	4
Obrázek 2: Cenová mapa Asociace realitních kancelářů	5
Obrázek 3: Cenová mapa Státního fondu rozvoje bydlení	7
Obrázek 4: National home prices page	8
Obrázek 5: Náhled tabulky prodeje bytů	15
Obrázek 6: Náhled tabulky obcí	17
Obrázek 7: Náhled tabulky ulic	17
Obrázek 8: Entitně relační model databáze	18
Obrázek 9: Crow's foot data model	18
Obrázek 10: Funkcionalita stahovacího skriptu	19
Obrázek 11: Cena bytů v krajích v roce 2013	30
Obrázek 12: Cena bytů v okresech v roce 2013	31
Obrázek 13: Cena bytů v obcích v roce 2013	32
Obrázek 14: Ceny bytů dle ulic v roce 2013 v Ostravě Porubě	33
Obrázek 15: Zobrazení pomocí interpolace	34
Obrázek 16: Model zpracování propojení dat	35
Obrázek 17: Obce s duplicitními názvy	36
Obrázek 18: Výsledek atributového dotazu	40
Obrázek 19: Výsledná mapová aplikace	41
Obrázek 20: Příklad změny hodnot legendy pronájmu bytů při změně měřítka	42
Obrázek 21: Vyhledávání adres a informace o aplikaci	43

Seznam tabulek

Tabulka 1: Shrnutí cenové mapy ze serveru bezrealitky.cz	5
Tabulka 2: Shrnutí cenové mapy Asociace realitních kancelářů	6
Tabulka 3: Shrnutí cenové mapy Státního fondu rozvoje bydlení	8
Tabulka 4: Shrnutí cenové mapy ze serveru Trulia	9
Tabulka 5: Ceny nemovitostí dle krajů	27
Tabulka 6: Prodej bytů ve vybraných obcích	28
Tabulka 7: Porovnání cen bytů ve vybraných lokalitách	44
Tabulka 8: Porovnání cen domů ve vybraných lokalitách	44
Tabulka 9: Shrnutí cenové mapy vyvinuté při této diplomové práci	44

Seznam příloh na CD

DB výstupy	- tabulky výstupů z databáze ve formátu xls
DB záloha	- záloha databáze k lednu 2014
Geoserver workspace	- pracovní prostor s veškerým nastavením, styly a šablonami
Skripty stahování	- skripty v jazyce php ke stahování inzerce z RSS kanálu
SQL Dotazy	- dotazy pro filtraci, agregaci a selekci inzerce z databáze
Web	- zdrojové soubory webových stránek včetně OpenLayers knihoven
Zpracování ArcMap	- projekt, toolbox, vstupní a výstupní vrstvy automatizovaného zpracování