

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

HODNOCENÍ KVALITY PROSTOROVÝCH DAT ZÍSKANÝCH CROWDSOURCINGEM

Magisterská práce

Bc. Jakub VRKOČ

Vedoucí práce RNDr. Jan BRUS, Ph.D.

Olomouc 2016
Geoinformatika

ANOTACE

Diplomová práce je zaměřena na hodnocení kvality prostorových dat získaných crowdsourcingem. Hlavním cílem bylo zhodnotit možnosti, jak lze kvalitu takto získaných prostorových dat zajistit. V rámci dílčích cílů teoretické části byla vypracována rešerše v oblasti crowdsourcingu, kvality prostorových dat, nejběžnějších standardů a aplikací využívající princip crowdsourcingu. Praktická část diplomové práce se skládá ze dvou částí. V první části byl navržen nástroj pro hodnocení polohové přesnosti založený na pravděpodobnostním rastru výskytu černých skládek pro území hlavního města Prahy. Pro tvorbu nástroje poskytl data IPR Praha a současně byla data získána z aplikace ZmapujTo. V obou případech se jedná o OpenData. Příprava dat byla provedena v softwaru ArcGIS for Desktop. Výpočet rastru byl realizován pomocí modelovacího softwaru Maxent. Výstupem je rastr reprezentující výskyt černých skládek v hlavním městě Praha. Pro zvýšení přesnosti bylo přistoupeno k optimalizaci. Optimalizace byla provedena expertním výběrem typů využití území s minimální pravděpodobností výskytu černých skládek s cílem snížit uvnitř této funkční plochy pravděpodobnost výskytu. Optimalizovaný rastr byl zhlazen pomocí filtru Focal Statistics. Výsledný pravděpodobnostní rastr byl testován na expertních a současně reálných datech s polohovou přesností dosahující 90 %. Druhá část práce obsahuje návrh konceptu pro komplexní hodnocení kvality prostorových dat. Koncept navrhuje možnosti zajištění jak polohové, tak atributové a časové přesnosti. V případě praktické realizace navrhovaného konceptu by vznikl poloautomatický nástroj pro komplexní hodnocení kvality prostorových dat v rámci aplikace ZmapujTo. Realizací konceptu by došlo k výraznému ušetření času administrátora, který v současnosti provádí kontrolu jednotlivých reportů manuálně. Výstupy práce budou poskytnuty IPR Praha. Úspěšným propojením crowdsourcingu a GIS vzniká zcela nová oblast geoinformatiky. V tomto spojení má geoinformatika značný potenciál pro budoucí využití.

KLÍČOVÁ SLOVA

crowdsourcing; kvalita dat; Maxent; ZmapujTo; černá skládka

Počet stran práce: 71

Počet příloh: 6 (z toho 5 vázaných a 2 volné)

ANOTATION

The diploma thesis is aimed at the quality evaluation of spatial data obtained by crowdsourcing. The main goal was to consider the possibilities of ensuring the spatial data quality. The theoretical part of the diploma thesis focuses on the research in the field of crowdsourcing, spatial data quality, the most common standards and applications using the crowdsourcing principle. The practical part is divided into two parts. The first one proposes a tool to evaluate spatial accuracy based on the probability grid of illegal disposal sites occurrence in the district of Prague. The data necessary for the tool creation was provided by IPR Praha as well as by the application ZmapujTo which are OpenData in both cases. The spatial data preparation was realized in ArcGIS for Desktop software. The grid computing was created by the help of Maxent software. The output is a grid representing the occurrence of illegal dumps in the capital city of Prague. Optimization was applied to increase the accuracy and it was performed by expert selection of land use types with minimal probability of illegal dumps occurrence keeping the objective to lower the probability of occurrence within the selected functional area. The optimized bitmap was smoothed using the Focal statistics filter. The resulting grid of probability was tested in both expert and real data with the accuracy achieving 90%. The second part deals with a concept draft for a complex evaluation of spatial data quality. The concept suggests possibilities of securing the spatial, attributive and time accuracy. In case of the proposed concept practical realization there would be created a semiautomatic tool for a complex evaluation of spatial data quality within the ZmapujTo application. The concept realization would significantly save time of the administrator who currently examines individual reports manually. The thesis output will be provided to IPR Praha. The successful connection between the crowdsourcing and GIS results in a whole new field of geoinformatics which has a significant capacity in such connection for the future use.

KEYWORDS

crowdsourcing; data quality; Maxent; ZmapujTo; illegal dump

Number of pages: 71

Number of appendixes: 7

Prohlašuji, že

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.

- jsem si vědom, 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 Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, bakalářskou/diplomovou práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk bakalářské/diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne

Bc. Jakub VRKOČ

Poděkování

Děkuji vedoucímu práce RNDr. Janu Brusovi, Ph.D. za cenné rady, podněty a připomínky při vypracování diplomové práce. Dále děkuji konzultantu RNDr. Miroslavu Kubáskovi, Ph.D., autoru aplikace ZmapujTo, za poskytnutí dat a možnost ověření konceptu hodnocení polohové přesnosti dat na reálných datech. Za poskytnutá podkladová data děkuji Institutu plánování a rozvoje hlavního města Prahy. V neposlední řadě děkuji své rodině a přítelkyni Míše za podporu v průběhu celého studia.

ZADÁNÍ - VLOŽIT

OBSAH

SEZNAM POUŽITÝCH ZKRATEK	X
ÚVOD	X
1 CÍLE PRÁCE.....	10
2 METODY A POSTUPY ZPRACOVÁNÍ.....	11
2.1 Použitá data	11
2.2 Použité programy	12
2.3 Postup zpracování.....	14
3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	15
3.1 Crowdsourcing	15
3.2 Kvalita prostorových dat	15
3.2.1 Standardy kvality prostorových dat.....	16
3.2.2 Elementy kvality prostorových dat	18
3.3 Kvalita crowdsourcingových dat	20
3.4 Využití crowdsourcingu v praxi	29
3.4.1 ZmapujTo	35
4 VLASTNÍ ŘEŠENÍ	40
4.1 Příprava vstupních dat.....	40
4.1.1 Úprava dat ze ZmapujTo.....	40
4.1.2 Příprava vstupních rastrů.....	43
4.2 Výpočet v softwaru Maxent	44
4.3 Testování a optimalizace	46
4.3.1 Testování I	47
4.3.2 Optimalizace I	48
4.3.3 Optimalizace II	49
5 VÝSLEDKY	51
5.1 Přesnost optimalizovaného rastru	51
5.2 Návrh konceptu pro komplexní hodnocení kvality hlášení v rámci aplikace ZmapujTo.....	54
6 DISKUZE	57
7 ZÁVĚR	59
POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE	
PŘÍLOHY	

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
AOPK ČR	Agentura ochrany přírody a krajiny ČR
ASCII	American Standard Code for Information Interchange
CC	Creative Commons
CDG	Crowdsourced Geospatial Data
CSV	Comma-separated values
ČÚZK	Český úřad zeměměřičský a katastrální
GIS	geografický informační systém
HOT	Humanitarian OpenStreetMap Team
INSPIRE	INfrastructure for SPatial InfoRmation in Europe
ISO	International Organization for Standardization
ISO/TC 211	ISO/TC 211 Geographic information/Geomatics
IPR Praha	Institut plánování a rozvoje hl. m. Prahy
KGI	Katedra geoinformatiky
OA	Open Address
ODbL	Open Database License
OGC	Open Geospatial Consortium
OSM	Open Street Map
RÚIAN	Registr uzemní identifikace, adres a nemovitostí
S-JTSK	Souřadnicový systém jednotné trigonometrické sítě katastrální
UGC	User Generated Geographic Content
ÚHUL	Ústav pro hospodářskou úpravu lesa
VGI	Volunteered Geographic Information
WGS-84	World Geodetic System 1984

ÚVOD

V současné době se využívání mobilních zařízení s připojením k internetu stalo zcela běžnou záležitostí. Lidé využívají mobilní aplikace jako nedílnou součást svého života. Mállokterý uživatel však vidí za mobilní aplikací zdroj prostorových informací. Specifickým typem aplikací jsou aplikace pracující na principu crowdsourcingu. Crowdsourcingem se rozumí způsob spolupráce komunity na daném projektu. Crowdsourcing funguje na principu dobrovolnosti.

U běžných prostorových dat je problematika kvality prakticky vyřešena. Jsou využívány standardy, rámce a postupy pro vytváření, respektive sběr prostorových dat. Z pohledu crowdsourcingu je zajišťování kvality takto vytvářených dat dosti problematické. Důvodem je především absence fáze standardizace v procesu vytváření dat. Crowdsourcing představuje pro GIS obrovský zdroj prostorových dat, které lze dále analyzovat. Velkým úskalím každého projektu realizovaného v geoinformatice je nákladnost na pořízení dat. V případě využití crowdsourcingu jsou data pořizována s minimálními náklady. Úspěšným propojením crowdsourcingu a geoinformatiky vzniká zcela nová oblast, které výzkum do současné doby nevěnoval patřičnou pozornost.

Na principu crowdsourcingu pracuje také aplikace ZmapujTo. V současné době probíhá kontrola jednotlivých reportů manuálně, což je časově dosti náročné. V rámci praktické části diplomové práce byl navržen a otestován nástroj pro hodnocení polohové přesnosti. Tento nástroj je součástí teoretického návrhu konceptu pro komplexní hodnocení kvality prostorových dat v rámci aplikace ZmapujTo. Praktickou realizací tohoto konceptu by došlo k výrazné úspoře času administrátora při kontrole jednotlivých reportů.



1 CÍLE PRÁCE

Cílem magisterské práce je zhodnotit možnosti hodnocení kvality prostorových dat získaných crowdsourcinglem. V úvodní části je provedena rešerše současných zdrojů dat získávaných crowdsourcinglem se zaměřením na současné metody sběru dat, standardy, rámce a možnosti ověřování kvality takto získaných dat. Současně jsou popsány nejběžnější aplikace využívající princip crowdsourcingu.

Praktická část diplomové práce je sestavena ze dvou podtémat. První část je zaměřena na tvorbu, optimalizaci a ověření distribučního modelu pro hodnocení pravděpodobnosti výskytu černých skládek na území Hlavního města Prahy. Tento model ověřuje prostorovou kvalitu bodových prvků – v tomto případě ilegálních skládek. Samotný model je ověřen na reálných datech, které poskytuje aplikace ZmapujTo. Výsledný výstup v podobě pravděpodobnostní mapy bude poskytnut IPR Praha.

V druhé části je navržen teoretický koncept pro zajištění komplexní kvality prostorových dat vytvářených v aplikaci ZmapujTo (polohová, atributová, časová složka). Takto navržený koncept je detailně popsán a vyjádřen v podobě schématického diagramu.

2 METODY A POSTUPY ZPRACOVÁNÍ

Použité metody, data, softwary a postupy zpracování byly po celou dobu realizace práce cíleně voleny pro optimální tvorbu predikčního modelu. Dílčí části této kapitoly jsou uvedeny níže strukturovanou a přehlednou formou.

2.1 Použitá data

Otevřená data neboli „OpenData“ jsou informace a data bezplatně a volně dostupná na internetu, pro jejichž využití nejsou kladeny legislativní či technické překážky. Formát a struktura otevřených dat umožňují jejich strojové zpracování, k němuž jejich vydavatel poskytuje právní svolení. Díky tomu pak mohou být data dále volně zpracovávána a využívána pro výzkum, tvorbu analýz, ale i ke komerčním účelům pro tvorbu internetových aplikací, čímž se zvyšuje jejich ekonomická hodnota a rozsah využití¹.

Pro účely diplomové práce bylo za oblast zájmu zvoleno hlavní město Praha. Toto území bylo vybráno záměrně, protože hl. m. Praha poskytuje vybraná prostorová data ve formě otevřených dat. Geografická data v otevřených datových formátech a webových službách poskytuje „*Institut plánování a rozvoje Hlavního města Prahy – IPR Praha*“ na svém geoportále Praha (<http://www.geoportalpraha.cz>).

Data jsou dostupná pod licencí CC BY-NC-ND 3.0 CZ (*Creative Commons Attribution Non Commercial No Derivates 3.0 Czech Republic*). Pod touto licencí lze data sdílet – to znamená rozmnožovat a distribuovat materiál prostřednictvím jakéhokoli média v jakémkoliv formátu. Poskytovatel licence nemůže tato oprávnění odvolat do té doby, dokud jsou dodržovány licenční podmínky. Je povinností uvést autorství, poskytnout s dílem odkaz na licenci a vyznačit provedené změny. Je zakázáno dílo užívat pro komerční účely. Pokud bude dílo zpracováno, doplněno nebo jinak změněno, není možno takovéto dílo dále šířit².

Data jsou publikována ve vektorových formátech Esri shapefile, GeoJSON, GML, DXF a rastrových formátech TIFF a JPG. Mimo rastrová data jsou všechny vektorové vrstvy publikovány v souřadnicových systémech S-JTSK a WGS-84. Vysokou kvalitu dat zaručují také podrobné metadatové záznamy.

Pro potřeby práce byly využity dvě datové vrstvy poskytnuté IPR Praha:

- Plán využití ploch – funkční plochy (územní plán)
 - souřadnicový systém WGS-84, prostorové rozlišení 1 : 5 000
- Současný stav využití území
 - souřadnicový systém WGS-84, prostorové rozlišení 1 : 10 000

Stejně jako IPR Praha, poskytuje prostorová data jako otevřená aplikace ZmapujTo. Data podléhají licenci CC BY-NC-SA 4.0 (*Creative Commons – Attribution-NonCommercial-ShareAlike 4.0 International*). Tato licence umožňuje data sdílet – rozmnožovat a distribuovat materiál prostřednictvím jakéhokoli média v jakémkoliv

¹ Pražská otevřená data: Co jsou otevřená data. Institut plánování a rozvoje hlavního města Prahy: Geografická data prahy na jednom místě [online]. Praha, 2015 [cit. 2016-03-30]. Dostupné z: <http://www.geoportalpraha.cz/cs/clanek/271/prazska-otevrena-data#.Vvu1d0dBQuM>

² CC BY-NC-ND 3.0 CZ: Uveďte původ-Neužívejte komerčně-Nezpracovávejte 3.0 Česká republika. Creative Commons: Keep the internet creative, free and open [online]. 2015 [cit. 2016-03-30]. Dostupné z: <https://creativecommons.org/licenses/by-nc-nd/3.0/cz/>

formátu. Dále upravit - remixovat, změnit a vyjít z původního díla. Poskytovatel licence nemůže tato oprávnění odvolat do té doby, dokud jsou dodržovány licenční podmínky. Je povinností uvést autorství, poskytnout s dílem odkaz na licenci a vyznačit provedené změny. Je zakázáno dílo užívat pro komerční účely. Pokud bude dílo zpracováno, doplněno nebo jinak změněno, není možno takovéto dílo dále šířit³. Datový formát je GeoJSON. Data jsou aktualizována ke stažení vždy o jeden den zpět na „datahube.io“.

Prostorová data uliční sítě Hl. m. Prahy byla získána při použití „*Veřejného dálkového přístupu k RÚIAN*“. Tato data nebyla použita jako vstupní proměnné a byla využita až následněm procesu optimalizace výsledného modelu.

2.2 Použité programy

ArcGIS 10.x for Desktop

ArcMap je centrální aplikace ArcGIS for Desktop, která slouží pro všechny mapové úlohy včetně kartografie, prostorových analýz a editace dat. Obsahuje většinu nástrojů, které umožňují data zpracovat, připravit pro analýzu, získat nové informace a výsledky zobrazit, popřípadě z nich vytvořit mapu nebo je odeslat na server⁴.

Pro zpracování a přípravu dat bylo nutno pracovat s ArcGIS for Desktop 10.1 nebo 10.2 s licenci „*Standard* nebo *Advanced*“ umožňující využití extenze *Spatial Analyst*. ArcGIS for Desktop 10.3 nebyl pro přípravu dat vhodný, protože některé funkce SDM toolboxu (uveden níže) nejsou plně funkční. Nativním formátem je Esri shapefile.

Pro zpracování dat do požadované formy využitelné v modelovacím softwaru Maxent bylo využito několika toolboxů (panel nástrojů):

- SDM toolbox v1.1c

Jedná se o toolbox pro ArcGIS for Desktop, který je založený na skriptovacím jazyce python. Slouží pro prostorové studie v oblastech ekologie, evoluce a genetiky. Skládá se ze 71 python-skriptů pro automatizaci výpočetních procesů v ArcGIS. Součástí toolboxu je možnost výpočtu maximální entropie tzv. Maxent neboli „*Maximum Entropy*“. Výstupem výpočtu je pravděpodobnost výskytu daného jevu v geografickém území⁵. Nástroj byl při práci využíván pro zpracování dat, které následně sloužily jako vstup do modelovacího softwaru.

- Prepare Rasters for Maxent Tool for ArcGIS 10.1

Toolbox umožňující přípravu dat pro vstup do modelovacího softwaru Maxent. Nástroj byl vytvořen pro verzi ArcGIS for Desktop 10.1. Základní verze je funkční při licenci „*Advanced*“. Současně je vyžadována extenze „*Spatial Analyst*“⁶.

³ CC BY-NC-SA 4.0: Uveďte původ-Neužívejte dílo komerčně-Zachovejte licenci 4.0 Mezinárodní. Creative Commons: Keep the internet creative, free and open [online]. 2015 [cit. 2016-03-30]. Dostupné z: <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.cs>

⁴ ArcGIS for Desktop: GIS začíná zde. ARCDATA PRAHA [online]. 2016 [cit. 2016-03-30]. Dostupné z: <https://www.arcdata.cz/produkty/arcgis/desktopovy-gis/arcmap>

⁵ SDM toolbox: Home [online]. Duke University, USA, 2015 [cit. 2016-03-30]. Dostupné z: <http://sdmtoolbox.org/>

⁶ Prepare Rasters for Maxent Tool for ArcGIS 10.1 [online]. 2015 [cit. 2016-03-30]. Dostupné z: <http://www.arcgis.com/home/item.html?id=11bf7e689c92413f8d31933b3e1f56b1>

- VFR Import Basic 10.2.11

Nástroj v podobě toolboxu určený pro převod dat z výměnného formátu RÚIAN (formát XML) do geodatabáze Esri. Balíček je dostupný volně ke stažení na webových stránkách společnosti ARCDATA PRAHA. Jako vstup slouží soubor formátu VFR, jehož stažení umožňuje webová aplikace „Veřejný dálkový přístup k datům RÚIAN“, dostupná na webových stránkách ČÚZK.

QGIS 2.6.0 - Brighton

QGIS je svobodný, multiplatformní, geografický informační systém. Do verze 2.0 býval označován také jako Quantum GIS. QGIS umožňuje prohlížení, tvorbu, editaci vektorových i rastrových prostorových dat, zpracování GPS dat a tvorbu mapových výstupů. V současnosti je nejnovější verzí QGIS 2.14 Essen. Výhodou tohoto softwaru je, že se jedná o open-source řešení⁷.

Maxent 3.3.3k

Software založený na výpočet maximální entropie pro uživatelem zvolené druhy. Tento software volí jako vstup environmentální proměnné (nadmořská výška, srážky a podobně) a geografický výskyt daného jevu. Výstupem je nejčastěji ASCII soubor (nemusí však být vždy) reprezentující pravděpodobnost výskytu jevu ve zvolené lokalitě.

Pro práci se softwarem Maxent v operačním systému Windows 7 je nutný plně funkční „Java runtime“ verze 1.4 a pozdější. Součástí balíčku je samotný software Maxent.jar (běžně spuštění), Maxent.bat (spouštění přes příkazový řádek), dále návod pro práci (tutoriál) a ukázková data. Softwarový balíček je volně dostupný ke stažení jak pro studijní a vědecké účely, tak pro komerční a neziskové účely. V případě komerčních aplikací požadují autoři dárcovství ze zisku, aby byl projekt Maxent dále tzv. *non-zisk*⁸.

Office 365

Office 365 je balík služeb a kancelářský software poskytovaný firmou Microsoft. Jedná se o reakci na zvyšující se popularitu cloudových služeb. Microsoft Office byl do té doby především samostatný kancelářský balík⁹. Pro potřeby diplomové práce byl využíván textový editor *Word*, tabulkový editor *Excel* a nástroj pro tvorbu prezentací *PowerPoint*.

Lucidpress a Lucidchart

Je webová aplikace umožňující upravovat vzhled, formát a design dokumentů, aniž by bylo nutné kupovat drahý software. Hlavní výhodou Lucidpressu je, že nabízí editační prostředí pro přípravu tištěných i digitálních publikací pro web nebo mobilní zařízení. Aplikace je propojená s *Google Drive* a umožňuje snadno importovat texty, obrázky i videoklipy například z *Youtube.com*¹⁰. Tato aplikace byla použita pro tvorbu posteru, který je jednou z povinných částí diplomové práce. Aplikace *Lucidchart* umožňuje jednoduše vytvářet vývojové diagramy a sdílet je mezi uživateli, například pro

⁷ QGIS: A Free and Open Source Geographic Information System [online]. 2016 [cit. 2016-03-30]. Dostupné z: <http://www.qgis.org/en/docs/index.html>

⁸ Phillips, S., Dudik, M. & Schapire, R., 2010, "Maxent Software, ver. 3.3.3e". [online]. 2015 [cit. 2016-03-30]. Dostupné z: <https://www.cs.princeton.edu/~schapire/maxent/>

⁹ Windows Central: What is Microsoft Office 365? [online]. Windows central - Richard Devine, 2015 [cit. 2016-03-30]. Dostupné z: <http://www.windowscentral.com/what-microsoft-office-365>

¹⁰ Lucidpress: AN ONLINE POSTER MAKER FOR EVERYONE [online]. Lucid Software Inc., 2015 [cit. 2016-03-31]. Dostupné z: <https://www.lucidpress.com/>

potřeby řízení projektů¹¹. Aplikace byla použita pro tvorbu rozhodovacích schémat, které jsou součástí diplomové práce.

PSPad 4.6.0

PSPad, celým názvem *PSPad Editor*, je freewarový textový editor a editor zdrojových kódů pro operační systém Microsoft Windows. Autorem programu je programátor Jan Fiala, první verze vyšla v roce 2001. V současnosti je *aktuální verze 4.6.0 (2007)*, která vyšla v říjnu 2015. Umožňuje pracovat v mnoha programovacích, skriptovacích a značkovacích jazycích¹². Tento software byl pro potřeby diplomové práce využit z toho důvodu, aby byl splněn požadavek plně funkčních a validních webových stránek, které jsou její součástí.

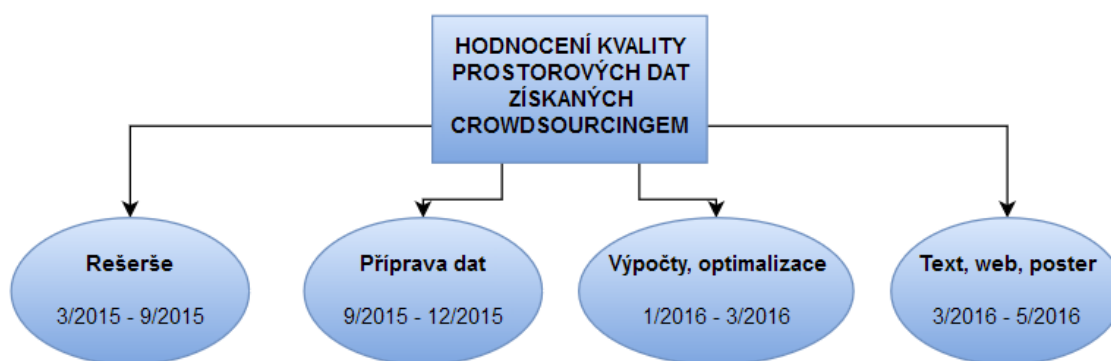
2.3 Postup zpracování

Samotný postup zpracování lze rozdělit do čtyř základních fází. První část se soustředila na seznámení se s crowdsourcingem obecně, s jeho problémy a možnými úskalími, výhodami, předpoklady a také s aplikacemi, které na principu crowdsourcingu pracují. Součástí teoretické fáze přípravy bylo studium způsobů zajištění kvality prostorových dat a nejběžněji používaných standardů. Výsledkem této části je kapitola 3. - „*Současný stav řešené problematiky*“.

Druhá fáze spočívala v získání prostorových dat potřebných pro praktickou část diplomové práce. Byl kontaktován RNDr. Miroslav Kubásek, Ph.D, který je autorem aplikace ZmapujTo. Další potřebná data byla získána z IPR Praha. Tato data bylo potřeba upravit do takové podoby, která by byla vhodná pro vstup do modelovacího softwaru Maxent.

Předposlední fázi tvořily dvě části. První část spočívala ve výpočtu a testování samotného pravděpodobnostního rastru v modelovacím softwaru Maxent. Na základě zjištěných výsledků došlo ve druhé části k *tzv. optimalizaci* výsledného rastru.

Finální fáze se skládala z vytvoření textové části diplomové práce. Jako součást práce jsou vytvořeny webové stránky a grafický výstup v podobě posteru. Graficky je postup zpracování znázorněn v obrázku 2.1.



Obrázek 2.1 Grafické znázornění postupu při zpracování diplomové práce.
autor: J. Vrkoč, 2016

¹¹ Lucidchart: Our flowchart maker works the way you do [online]. Lucid Software Inc., 2015 [cit. 2016-03-31]. Dostupné z: <https://www.lucidchart.com/>

¹² PSPad: freeware editor [online]. Slavkov u Brna: Jan Fiala, 2015 [cit. 2016-03-31]. Dostupné z: <http://www.pspad.com/cz/>

3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Tvorba, kvalita a využívání prostorových dat jsou moderním trendem dnešní doby. Prostorová data jsou součástí prakticky všech prací, studií, publikací a článků, ve kterých v určité míře figuruji geoinformační systémy a technologie. Samotnému hodnocení kvality dat však odborná veřejnost přikládá menší zájem. Prostorová data získaná crowdsourcingem a jejich kvalita je mnohem intenzivněji řešena v zahraničních studiích, přičemž českých článků a publikací je jen velmi málo.

3.1 Crowdsourcing

Slovo „*crowdsourcing*“ poprvé použil v červnu 2006 *Jeff Howe*. Crowdsourcing je slovní spřežka dvou anglických slov „*crowd = dav*“ a „*outsourcing = využívání subdodavatelských vztahů*“. Crowdsourcingem je chápáno najímání si davu či sdílení jednoho úkolu s větší skupinou lidí, většinou přes internet a často anonymně¹³. I když se jedná o nový termín, příklady crowdsourcingových projektů jsou známy již z počátku 18. století.

Metoda crowdsourcingu využívá efektivní komunikace, zájmu a chuti komunity vytvářet obsah. Komunita dokáže konkurovat profesionálům především skrze pořizovací cenu. Vlastní úloha je zadána komunitě a nikoliv jiné organizaci či firmě.

Výhoda crowdsourcingu s ohledem na sběr prostorových dat spočívá ve velkém množství nově vytvořených prostorových dat s minimálními náklady. Důvodem je spolupráce většího množství lidí, než by si daná organizace či firma mohla dovolit. Honorář může být v určitých případech vyplácen za výsledek daného projektu.

Mezi nevýhody crowdsourcingu patří možné navýšení nákladů potřebných k dokončení projektu. Současně se zvyšuje pravděpodobnost selhání projektu z důvodu nedostatku finanční motivace či malého počtu zúčastněných lidí. Dalším limitujícím faktorem může být skutečnost, že pouze nízké procento zúčastněných chce proniknout do dané problematiky. Z pohledu geoinformatiky je za hlavní nevýhodu označováno obtížné hodnocení kvality takto získaných prostorových dat.

Mobilní zařízení (smartphony, tablety a podobně) jsou v současnosti nejčastěji využívaným prostředkem, pomocí kterých je realizován rychlý, efektivní a levný sběr prostorových dat. Aplikací využívajících principu crowdsourcingu existuje nepřeberné množství. Mezi evropsky nejuznávanější patří aplikace jako *OpenStreetMap*, *Waze* a *Flickr*.

3.2 Kvalita prostorových dat

Pro stanovení a definování kvality prostorových dat je důležité pochopit, co slovo „*kvalita*“ znamená. Význam tohoto pojmu prošel a stále prochází vývojem a proměnou. Od počátku geoinformatiky je kvalita prostorových dat její klíčovou subdisciplínou. V současnosti jsou odbornou veřejností přijímány dva základní přístupy pro hodnocení „*kvality prostorových dat*“ nebo „*nejistoty prostorových dat*“.

¹³ HOWE, J. Crowdsourcing:: How the power of the crowd is driving the future of business [online]. Crown business, 2008, 320 s. [cit. 2016-20-03].

Jako nejvhodnější definice kvality dat se ukázala ta, kterou použil ve své disertační práci Brus¹⁴. Jedná se o každou jakost či vlastnost, která něčemu přísluší, nebo se k něčemu připojuje. Samotnou kvalitou je chápána *kvalita získané informace*. Z tohoto úhlu pohledu lze za kvalitní považovat takovou informaci, která je užitečná a použitelná pro svého „spotřebitele“, přičemž ve spojení s tradiční definicí informace to v geoinformatice znamená, že kvalita informací je závislá nejen na základních datech, ale především na uživateli samotném¹⁴.

Autor uvádí, že existují dva předpoklady pro kvalitní data. Zaprvé je potřeba dodržet požadavky uživatelů během produkce nebo transformace prostorových dat. Druhý aspekt odpovídá dokumentaci, která je poskytována uživatelům dat. Data s vysokou kvalitou je možné označit jako „přesná a důvěryhodná“. Celá subdisciplína naráží na nejednotnost používaných definic a nepřesnou terminologii. Informace o kvalitě prostorových dat jsou nezbytné pro jejich výběr a správné používání. Cílem poskytovatelů dat je jejich znovu použití¹⁴.

Z hlediska kvality crowdsourcingových dat existuje velmi tenká linie mezi použitelnými (kvalitními) a nepoužitelnými (nekvalitními) daty. U běžných dat probíhá v samotné části tvorby dat fáze standardizace. U crowdsourcingových dat je velmi obtížné určit kvalitu geodat bez fáze standardizace.

3.2.1 Standardy kvality prostorových dat

Problematika standardů, norem a profilů týkajících se kvality prostorových dat je značně rozsáhlá. Z hlediska mezinárodní platnosti vznikaly standardy hlavně v rámci „Mezinárodní organizace pro standardizaci - ISO“.

ISO standardy v procesu zajištění kvality geografických produktů

Pro geoinformatiku jsou, především z pohledu kvality prostorových dat, zásadní normy vzniklé pod záštitou ISO/TC 211. ISO/TC 211 je standardizační technická komise se zaměřením na geografickou informaci a geomatiku. Výsledkem práce komise je celá řada norem ISO19100 vytvořených za účelem definování, popisu a správy geografických informací¹⁵. Komise ISO/TC 211 vyvíjí standardy pro vyhledávání a používání geografických informací bez ohledu na platformu. Nezabývá se pouze řešením technických a sémantických otázek, ale také metadat, které mají na kvalitu prostorových dat podstatný vliv. V následující části je uveden výčet základních ISO norem, které jsou využity pro hodnocení kvality prostorových dat.

- Specifikace produktu – Data product specification (ISO 19131)

ISO 19131 vymezuje požadavky na specifikaci produktů geografických dat, založené na pojetí jiných norem z řady ISO 19100. Norma obsahuje informace o vytvoření specifikace datového produktu, terminologii, zkratkách, názvu datového produktu a zahrnuje informační popis datového produktu¹⁶.

¹⁴ BRUS, Jan. Vizualizace nejistoty v environmentálních studiích. Olomouc, 2013. disertační práce. 24-25s. UNIVERZITA PALACKÉHO V OLOMOUCI. Přírodovědecká fakulta

¹⁵ KRESSE, W., FADAIE, K. ISO standards for geographic information. Berlin; New York: Springer, 2004. ISBN 978-3-540-20130-4.

¹⁶ ISO. ISO 19131:2007 Geographic information - Data product specifications. 2007.

- Metadata (ISO 19115)

ISO 19115 definuje schéma, které je potřebné pro popis geografických informací a služeb. Poskytuje informace o identifikaci rozsahu, kvalitě dat, prostorovém a časovém schématu, územním rozhodnutí a distribuci geografických dat. Standard definuje povinné a podmíněné sekce metadat, metadat objektů a metadata prvků, minimální soubor metadat potřebných k definování metadatových aplikací v jejich plném rozsahu (tj. údaje o vzniku, určení vhodnosti dat pro jejich použití, přístup k datům a jejich přenos a využití digitálních dat). Umožňuje vytvářet i tzv. *profily* – upravený standard kompatibilní se svým výchozím standardem, který je upraven pro určité specifické podmínky (např. profesní)¹⁷.

- Měření kvality dat – Data quality measures (ISO 19138)

Norma definuje soubor opatření týkajících se kvality údajů a poskytnutí správných a standardizovaných metod pro měření kvality prostorových dat. Definuje seznam měření pro kvantifikaci prvků kvality podle ISO 19113. Tato norma specifikuje povinné elementy: název, prvek kvality, podprvek, definice a typ hodnoty kvality dat¹⁸.

- Kvalita dat – Data quality (ISO 19157)

Dle výše uvedených norem vyplývá, že pro úspěšné hodnocení kvality je zapotřebí mít o celé řadě ISO norem přehled. Na základě toho vznikla norma ISO 19157, která je harmonizována do jednoho nového celku¹⁹. Jedná se o nejnovější normu schválenou v květnu 2015. Na základě schválení normy ISO 19157 došlo ke zrušení tří níže uvedených norem, a to:

- Zásady jakosti – ISO 19113 – vydaná v červenci 2004
- Postupy hodnocení jakosti – ISO 19114 – vydaná v květnu 2005
- Kvalita – ISO 19157 – vyhlášena v srpnu 2014²⁰

Mimo ISO normy existují i další standardy, které lze do problematiky kvality prostorových dat zahrnout. Velmi často se však jedná o normy vytvářené také pomocí ISO norem; popřípadě se jedná pouze o ISO normy poupravené. Eventuálně mohly tyto standardy posloužit jako základní kámen pro vytvoření ISO normy.

OGC - Open Geospatial Consortium

Paralelně s prací komise ISO/TC211 vytvářelo soubor norem také OGC. Standardy OGC jsou však silně ovlivněny normami vytvořenými právě touto komisí. Hlavní přednost OGC je spočívá v prosazování formátu pro výměnu vektorových dat *Geography Markup Language (GML)*, specifikace pro služby *Web Map Server Interface (WMS)* a *Web Feature Services (WFS)*. Tyto implementace jsou také přebírány a zveřejňovány komisí ISO/TC 211 jako normy nebo technické specifikace v momentě, kdy se dostanou do dostatečného stádia realizace²¹.

¹⁷ ISO. ISO 19115:2003 Geographic information - Metadata. 2003b.

¹⁸ ISO. ISO/TS 19138:2006 Geographic information - Data quality measures. 2006.

¹⁹ ISO. ISO/DIS 19157 Geographic information - Data quality 2013.

²⁰ Věstník: Úřadu pro technickou normalizaci, metrologii a státní zkušebnictví [online]. 2015(5), 68 [cit. 2016-04-02]. Dostupné z: <http://www.technickenormy.cz/publicdoc/vestnik-05-15.pdf>

²¹ BRUS, Jan. Vizualizace nejistoty v environmentálních studiích. Olomouc, 2013. disertační práce 29s. UNIVERZITA PALACKÉHO V OLOMOUCI. Přírodovědecká fakulta

INSPIRE – Infrastructure for SPatial InfoRmation in Europe

INSPIRE vznikl na základě iniciativy Evropské komise a současně se jedná o směrnici Evropské komise a Rady. Cílem je vytvořit evropský legislativní rámec potřebný k vybudování evropské infrastruktury prostorových informací. Stanovuje obecná pravidla pro založení infrastruktury prostorových dat, zejména v oblasti environmentálních politik a životního prostředí.

K základním principům směrnice INSPIRE patří:

- sbírat, vytvářet a spravovat data jednou na co nejefektivnější úrovni,
- bezešvě kombinovat data z různých zdrojů a sdílet je mezi nejrůznějšími uživateli nebo aplikacemi,
- sbírat data na jedné úrovni a využívat je na jiných úrovních,
- poskytovat data za podmínek, které neomezí jejich rozsáhlé využívání,
- data a služby popsat metadaty pro jejich snadnější vyhledávání a zhodnocení²².

Nejdůležitějšími dokumenty z hlediska INSPIRE jsou:

- *Nařízení komise č. 1205/2008* (prosinec 2008) – týká se metadat,
- *Nařízení komise č. 1089/2010* (listopad 2010) – interoperabilita prostorových datových sad,
- *INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119* (červen 2010).

Výše uvedené standardy jsou všeobecně nejvíce používány a přijímány odbornou veřejností. Ve světě jsou užívány další standardy pro stanovení kvality prostorových dat jako:

- Standard pro digitální geoprostorová metadata - CSDGM (USA),
- ANZLIC Metadata profile - (Australian and New Zealand Information Council),
- National Transfer Format – NTF (Velká Británie).

Hans-Jörk Stark z Centre of Geoinformatics v Salzburku uvádí, že normy ISO 19100 jsou nejvhodnějším řešením pro hodnocení kvality prostorových dat získaných metodou crowdsourcingu²³. V současné době neexistuje standard, který by zaštiťoval crowdsourcingová geodata, popřípadě deklaroval určitý stupeň kvality takovýchto dat. Mnohem efektivnější než vytvářet standard nový je využít standardy stávající, popřípadě upravené pro potřeby takto vytvářených dat.

3.2.2 Elementy kvality prostorových dat

Všeobecně přijímaným faktem je, že kvalita prostorových dat je tzv. *multikomponentní*, a proto neexistuje pouze jedno pojetí kvality prostorových dat (příkladem může být velké množství standardů). Kvalitu prostorových dat lze hodnotit z pohledu jednotlivých elementů – jejich výčet je uveden v tabulce 3.1.

²² TRYHUBOVÁ, P. Principy INSPIRE a standardizace ve vazbě na data ZABAGED R. 2013. Disertační práce. České vysoké učení technické v Praze. Fakulta stavební, Katedra mapování a kartografie

²³ JÖRG STARK, Hans. *Quality assurance of crowdsourced geocoded address-data within OpenAddresses: Concepts and Implementation*. Centre for GeoInformatics, Salzburg University, 2010, 112-113s. Master Thesis. Centre for GeoInformatics, Salzburg University.

Tabulka č. 3.1 Elementy kvality prostorových dat²⁴

		Čas	Atribut	Měřítko	Vztah
Přesnost	Polohová přesnost	Časová přesnost	Tematická (atributová) přesnost	-	-
Preciznost	-	-	-	-	-
Spolehlivost	-	-	-	-	-
Rozlišení	Prostorové rozlišení	Časové rozlišení	Tematické rozlišení	-	-
Logická konzistence	Prostorová konzistence	Časová konzistence	Doménová konzistence	-	Topologická konzistence
Úplnost	-	-	-	Úplnost	-

Polohová přesnost

Polohová přesnost zahrnuje parametry přesnosti určení horizontální a vertikální polohy geoprvků v prostorové datové sadě. Tato přesnost určování nezávisí jen na použité metodě (např. GPS, fotogrammetrie, geodetické zaměření), ale také na získaných zkušenostech mapující osoby, respektive osoby tvořící datovou sadu. Kvalitu prostorových dat lze rozlišovat na relativní a absolutní. Jako příklad je možno uvést měření vzdálenosti mezi dvěma body na Zemi a měření vzdálenosti mezi totožnými body na mapě – tato vzdálenost musí být v rozmezí relativní přesnosti. Absolutní přesnost v poloze je ukazatelem toho, s jakou přesností jsou prostorové objekty umístěny na mapě s ohledem na jejich přesnou polohu na Zemi v absolutním referenčním rámci (např. WGS-84)²⁵. Možností jak zajistit kvalitu dat z hlediska polohové přesnosti se zabývá praktická část diplomové práce v kapitole 4.

Tematická (atributová) přesnost

Tematická přesnost vyhodnocuje spolehlivost, respektive přesnost hodnot prvků v datové sadě ve vztahu k jejich skutečnému významu v „reálném světě“. Součástí je také údaj posuzující, jak přesně atributy odpovídají klasifikační metodě (obvykle vyjádřeno v procentech). V mnoha případech se jedná o velmi subjektivní hodnocení při stanovování hodnoty atributu²⁵. Příkladem může být správné odlišení černé skládky – respektive rozlišení z hlediska velikosti, zda se jedná o skládku střední velikosti do 50 kg, nebo o skládku přesahující 50 kg).

Časová přesnost

VÚGTK (Výzkumný ústav geodetický, topografický a kartografický, v.i.i.) popisuje časovou přesnost jako parametr jakosti udávající přesnost časových aspektů geografických dat. Z hlediska časové přesnosti lze vymezit správnost a přesnost, pomocí kterých jsou uvedeny chyby v měření²⁶. Časová přesnost se týká informací

²⁴ SHI, W. Principles of modeling uncertainties in spatial data and spatial analyses. Boca Raton: CRC Press/Taylor & Francis, 2010. ISBN 978-14-20059-27-4

²⁵ BRUS, Jan. Vizualizace nejistoty v environmentálních studiích. Olomouc, 2013. disertační práce). 33s. UNIVERZITA PALACKÉHO V OLOMOUCI. Přírodovědecká fakulta

²⁶ PRAŽÁK, J.: Terminologický slovník zeměměřičství a katastru nemovitostí [online]. Zdičky Výzkumný ústav geodetický, topografický a kartografický, Terminologická komise Českého úřadu zeměměřického a katastrálního, 2005 - 2013, [cit. 12-12-2012]. Dostupné na WWW: <<http://www.vugtk.cz/slovník/index.php>>.

zaznamenávající datum pozorování, druh aktualizace a časový interval, po kterém jsou data platná. U některých typů prostorových dat má parametr „času“ zvláštní důraz, především u dat, která se neustále mění (např. současný stav využití území – dochází zde k neustálým změnám).

Výše uvedené elementy spadající pod přesnost (polohová, časová, tematická) nejsou však jedinými, které se využívají pro hodnocení kvality prostorových dat. Současně s nimi bývá hodnocena:

- sémantická přesnost,
- úplnost,
- logická konzistence,
- správnost,
- rozlišení,
- původ,
- účel,
- použitelnost,
- rodokmen,
- metakvalita,
- homogenita,
- spolehlivost.

Jednotlivé elementy detailněji popisuje v disertační práci Jan Brus²⁷. Pro crowdsourcingové data je z výše uvedených důležitá právě přesnost polohová a tematická, neboli atributová. Ostatní elementy nabývají na důležitosti až s ohledem na danou problematiku, typ mapování a způsob získávání prostorových dat crowdsourcingem. V praktické části této diplomové práce je vytvořen distribuční model pro hodnocení prostorové kvality bodových prvků – příkladem může být výskyt černých skládek, umístění billboardů podél významných komunikací a podobně.

3.3 Kvalita crowdsourcingových dat

Množství zahraničních studií zabývajících se prostorovými daty vytvořenými metodou crowdsourcingu vysoce převyšuje počet studií tuzemských. Hojně publikované studie a články jsou především v anglickém jazyce. Výhodou je široká škála záběru samotné problematiky, ať už se jedná o aplikace využívající crowdsourcing, o kvalitu takto získaných dat, nebo o jejich praktické využití v marketingu, žurnalistice, krizovém managementu a dalších formách využití geografických informačních systémů (GIS).

V roce 2010 publikovali Jeffrey Heer a Michael Bostock z University of Stanford článek s názvem „*Crowdsourcing graphical perception: using mechanical turk to assess visualization design*“²⁸. Autoři vyjadřují myšlenku, že geografická data popisují svět z různých úhlů pohledu pomocí přímého či nepřímého vyjádření zeměpisné polohy. Autor vyjadřuje znepokojení nad tím, že geografická data jsou společností zaměňována a používána pro jiné účely než pro ty, pro něž byla primárně vytvořena. Informace o

²⁷ BRUS, Jan. Vizualizace nejistoty v environmentálních studiích. Olomouc, 2013. disertační práce. 34-36s. UNIVERZITA PALACKÉHO V OLOMOUCI. Přírodovědecká fakulta

²⁸ HEER, Jeffrey a Michael BOSTOCK. *Crowdsourcing graphical perception:: using mechanical turk to assess visualization design* [online]. Stanford University, 2010, 1-5s. [cit. 2015-06-06]. ISBN 978-1-60558-9299. Dostupné z: <http://dl.acm.org/citation.cfm?id=1753357/>

kvalitě prostorových dat jsou nezbytné ke správnému výběru použitých dat pro následné analýzy. Je důležité uvědomit si, že v reálné situaci nejsou k dispozici prostorová data, která lze označit za perfektní. Tento předpoklad platí pro jakoukoliv oblast vědy včetně kartografie a geoinformačních systémů²⁹.

Nedokonalost dat (nepřesnost) má nicméně nezanedbatelný vliv na všechny procesy ve vědách využívající GIS. V posledních letech začaly tyto nepřesnosti zajímat specialisty pracující s prostorovými daty, kteří začali pracovat na možnostech jejich odstranění. Chyby a nepřesnosti jsou jedny z mnoha parametrů, které mají rozhodující vliv na kvalitu dat. V případě, že chyby již vznikají, je důležité uvědomit si a přesně popsat způsob, jak dané chyby vznikají. Popřípadě, jak je lze v dalším řešeném projektu odstranit. V praxi tyto aspekty podnítily geoinformační společnost k vytvoření velkého množství norem a standardů. Účelem je co možná nejvyšší opětovné využití geodat²⁹.

Součástí prostorových dat by vždy měl být dokument sestávající ze dvou částí. V první jsou popisovány požadavky uživatelů, které by měly být splněny v průběhu tvorby a transformací dat. Druhá část by měla exaktně popisovat kvalitu prostorových dat z hlediska různých aspektů. V oblasti crowdsourcových dat je však situace odlišná. Je důležité si uvědomit, že tento typ dat vytvářejí a shromažďují osoby, kterým chybí odborná kvalifikace či vzdělání potřebné pro tuto činnost. Data jsou sbírána za pomoci dobrovolnické činnosti. Za tvorbu metadatových záznamů však nikdo zodpovědnost nenese. U crowdsourcových dat však zpravidla chybí dříve uváděná dokumentace popisující kvalitu prostorových dat. Pro rozumnou a přesnou práci s těmito daty je naprosto nezbytné, aby probíhala kontrola kvality. Toto je základní předpoklad k tomu, aby mohla být data přijata pro práci²⁹.

U crowdsourcových dat existuje velmi tenká linie mezi použitelnými a nepoužitelnými daty z hlediska jejich kvality. U běžných dat probíhá v samotné části jejich tvorby fáze standardizace. U crowdsourcových dat je velmi obtížné určit všeobecně přijímaný způsob, na základě kterého by byla stanovena kvalita geodat získaných bez fáze standardizace. Jedním z možných řešení, jak zaplnit mezeru mezi použitelností dat a náročnou kontrolou kvality je využití grafického znázornění kvality dat – tzv. „*geovisual*“. Jedná se o grafické vyjádření prvků kvality prostorových dat. Vzhledem ke složitosti problému s hodnocením kvality stanovuje tato metodika různé faktory, které mohou mít vliv na kvalitu geodat, popřípadě na jejich nepřesnosti. V současné době se studie zabývají především využitím vizualizačních technik, které jsou používány pro vyjádření nejistoty prostorových dat²⁹.

Další diplomovou prací v roce 2010 je „*Quality assurance of crowdsourced geocoded adress-data within OpenAdresses*“ od Hanse-Jörg Starka z Centre of Geoinformatics v Salzburku. Autor v závěru práce popisuje sedm klíčových zjištění, kterých dosáhl při vypracovávání studie. Těchto sedm dílčích zjištění je uvedeno níže³⁰:

- 1) Norma ISO/TC 211 19100 se ukázala být vhodná pro hodnocení kvality crowdsourcových dat.

²⁹ HEER, Jeffrey a Michael BOSTOCK. *Crowdsourcing graphical perception:: using mechanical turk to assess visualization design* [online]. Stanford University, 2010, 1-5s. [cit. 2015-06-06]. ISBN 978-1-60558-929-9. Dostupné z: <http://dl.acm.org/citation.cfm?id=1753357/>

³⁰ JÖRG STARK, Hans. *Quality assurance of crowdsourced geocoded address-data within OpenAddresses: Concepts and Implementation*. Centre for GeoInformatics, Salzburg University, 2010, 112-113s. Master Thesis. Centre for GeoInformatics, Salzburg University.

- 2) OWMS lze úspěšně integrovat do real-time hodnocení kvality pro „OpenAdresses – OA“ data.
- 3) Vyhodnocení správnosti atributů adresy jsou správné pouze ze 75 %. To je způsobeno přísně nastaveným binárním algoritmem, který byl pro tuto práci použit.
- 4) Poziční přesnost je obtížné posoudit, protože chyba vzdálenosti mezi skutečnou polohou a OWMS interpolovaným umístěním se může výrazně lišit. I když malé poziční chyby nemusí být zjištěny, tak hrubé chyby jsou identifikovány.
- 5) Online přehled real-time hodnocení kvality OA dat je užitečné a praktické, protože umožňuje real-time úpravy chybných údajů v databázi prostřednictvím grafického uživatelského rozhraní OA.
- 6) Pro vizuální dojem jsou využívány Google Maps. Takto jsou vyhodnocovány záznamy s malými pozičními chybami.
- 7) Tato práce potvrzuje, že méně přesné referenční datasety mohou pomoci v posuzování lepších datasetů v tom smyslu, že jsou ukazatelem zejména hrubých chyb³¹.

V listopadu roku 2012 byla vydána doposud nepřekonaná rozsáhlá publikace „*Crowdsourced Geospatial Data*“ s podtitulem „*A report on the emerging phenomena of crowdsourced and user-generated geospatial data*“. Autorem je Matthew T. Rice a kol. z katedry geografie a geoinformačních věd Univerzity George Masona. Autoři uvádějí, že rozvíjející se fenomén crowdsourcových prostorových geodat je důležitým trendem pro společnost lidí zabývajících se problematikou prostorových dat. Ve své práci se věnují především způsobu využití metody crowdsourcingu, na základě které probíhá generování, shromažďování, uchovávání a prezentování získaných geodat. Dále je zde řešeno zapojení koncových uživatelů do společnosti věnující se crowdsourcingu. Mnoho z těchto uživatelů není v oblasti geografie a geoinformačních systému vzděláno; proto je také autoři označují pojmem „neogeografové“. Ti společnosti zabývajících se crowdsourcingem přispívají především sběrem geoprostorových dat, vývojem geoprostorových aplikací a dalšími mediálními aktivitami. Ve snaze pochopit souvislost mezi kvalitou dat a neogeografií je nutné definovat tzv. *VGI (Vooluntereed Geographic Information)* jako součást *UGC (User Generated Geographic Content)*³².

Crowdsourcová prostorová data (CGD) mají levnou výrobní cenu, což umožňuje, aby byly generovány pro velké plochy za pomoci dobrovolné činnosti účastníků. Data mohou být vytvořena velmi rychle, jak lze doložit na četných případech mimořádných událostí a katastrof, při kterých byla tímto způsobem data vytvářena. Crowdsourcová data bývají zpravidla generována prostřednictvím nástrojů a v současné době také mobilními aplikacemi, které jsou open-source. Posledním, ale naprosto důležitým předpokladem

³¹ JÖRG STARK, Hans. *Quality assurance of crowdsourced geocoded address-data within OpenAdresses: Concepts and Implementation*. Centre for GeoInformatics, Salzburg University, 2010, 112-113s. Master Thesis. Centre for GeoInformatics, Salzburg University.

³² RICE, Matthew T a kol. *Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data* [online]. 1. George Mason University, 2012, 124-126s.[cit. 2015-06-05]. Dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf>

pro kvalitní tvorbu dat jsou zeměpisné znalosti přispěvatelů. Tento předpoklad je naprosto zásadní pro tvorbu prostorových dat získaných crowdsourcingem³³.

Publikace se skládá z několika důležitých částí, mimo jiné také z kapitoly o projektu OSM (*OpenStreetMap*), který má za cíl produkovat celosvětová a editovatelná data zdarma. Původ OSM lze dohledat v open-source hnutí ve Velké Británii především jako reakci na přísné licenční politiky společnosti „Ordnance Survey“, která se zabývá tvorbou geografických dat ve Velké Británii³³.

Další důležitou částí je také hodnocení kvality prostorových dat, která jsou vytvořena metodou crowdsourcingu. Způsobem hodnocení kvality se zabývá přední výzkumník z oblasti geodat Michael F. Goodchild, který se podílí na tradičním hodnocení přesnosti už několik desetiletí. Za nejnovější přístup označují autoři tzv. Linusův zákon, který je založen na pravidlech třídění geografických dat již v průběhu jejich tvorby. Nejdůležitějším obecným nástrojem pro hodnocení kvality geoprostorových dat je plné využití potenciálu metadat. Je vytvořeno mnoho vynikajících metadatových standardů pro geoprostorová data. Tyto standardy by však měly být plně implementovány v rámci crowdsourcového vytváření dat. V případě požadavků na využití crowdsourcových prostorových dat v projektech, popřípadě aplikacích, je nutné provést:

- a) kontrolu metadat (pokud existují),
- b) vizualizaci nejistoty,
- c) srovnání crowdsourcových dat s referenčními daty,
- d) znalecké posudky a hodnocení CGD,
- e) rizika a výhody pro použití CGD,
- f) způsob tvorby CGD³³.

V určitých typech projektů vyžadujících vyšší časovou náročnost mohou převážít výše uvedené výhody a rizika spojená s kvalitou tvorby těchto dat (poziční přesnost). Rizika jsou však v tomto případě akceptovatelná. Typickým příkladem jsou aplikace mapující živelné katastrofy (zemětřesení, povodně, požáry a podobně).

Za další důležitý předpoklad pro optimální tvorbu prostorových crowdsourcových dat autoři označují efektivní rozvoj a zapojení co možná nejrozsáhlejší komunity uživatelů. Zapojení velkého počtu uživatelů je základním předpokladem pro dlouhodobou životnost CGD projektů, jako je OSM. Významná pozornost musí být věnována povzbuzující a motivující činnosti (CDG spoléhá na dobrovolné pracovní síly). Dle autorů je vhodné použít určitý typ opatření ke kontrole kvality, což může pomoci ke zlepšení kvality příspěvků. Příkladem může být požadavek na přihlášení uživatele, aby mu byla umožněna editace prvků v mapě. Toto „omezení“ však může vézt i k opačnému efektu- může například způsobovat nižší míru účasti uživatelů na projektu. V případě, že je prostředí, ve kterém dochází k tvorbě CDG pro uživatele otevřené, je však také současně otevřené chybám a vandalskému chování ze strany uživatelů. Velké společnosti, jakou je OpenStreetMap nebo Wikipedie, vyvinuly analytické nástroje pro detekci škodlivých příspěvků. Tyto nástroje mají za úkol zvyšovat kvalitu příspěvků³³.

V předposlední části se autoři zabývají licencemi, pod kterými jsou šířena prostorová crowdsourcingová data. Překvapivě, otevřené licenční podmínky mohou omezovat veřejné použití prostorových dat získaných crowdsourcingem. Například OSM data jsou

³³ RICE, Matthew T a kol. *Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data* [online]. 1. George Mason University, 2012, 20-60s.[cit. 2015-06-05]. Dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf>

šířená bez omezení použití. Řady licencí byly vyvinuty prostřednictvím licencí „Creative Commons“ a „Open Knowledge Foundation“³⁴.

Komplexně autoři hodnotí crowdsourcingová prostorová data jako jednu z nejlepších myšlenek pro tvorbu dat. Odborná veřejnost začíná uznávat tento typ geodat v plném rozsahu. Stále si však kladou otázky, o jak spolehlivý a důvěryhodný datový zdroj se jedná. Uvedená publikace poskytuje ucelený pohled na problematiku prostorových dat získaných crowdsourcingem. Je určena především jednotlivcům a organizacím zabývajících se příslušnou problematikou ať už v oblasti tvorby dat, či jejich využíváním³⁴.

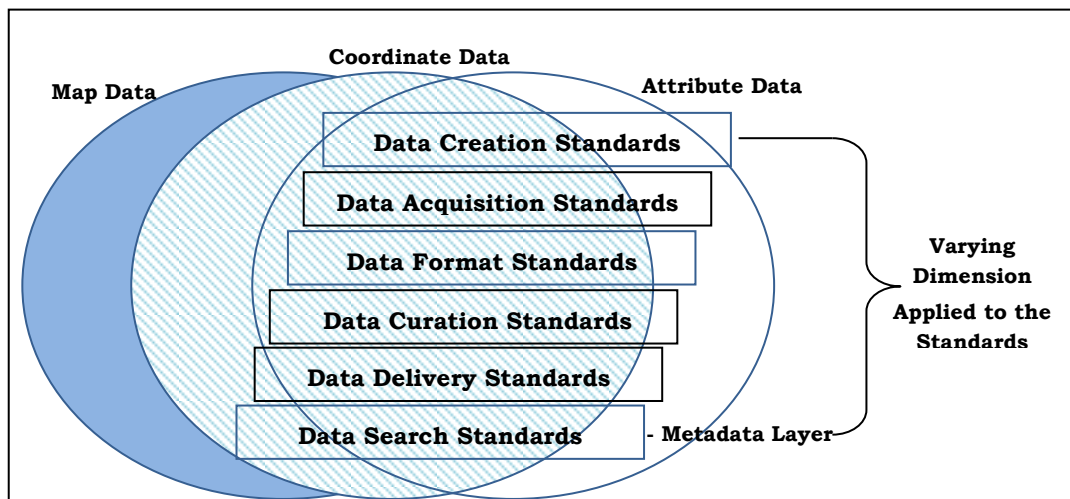
V roce 2012 publikoval článek Jingfeng Xia z Indiana University v Indianapolis s názvem „Metrics to Measure Open Geospatial Data Quality“. Autor se v článku zaměřuje na jedinečnost a složitost geoprostorových dat, především pak na hodnocení kvality. Touto vyzývavou otázkou se zabývají jak poskytovatelé dat a datových služeb, tak analytici, kteří s vytvořenými daty dále pracují. Autor popisuje dva základní přístupy pro hodnocení kvality. Jedná se o kvantitativní metriku založenou na objektivním měření a kvalitativní metriku založenou na subjektivním měření. Tyto přístupy jsou nezbytné pro hodnocení kvality geoprostorových dat. Optimálním řešením je propojení obou přístupů a následné provedení tzv. *multi-level* hodnocení. Hodnocení kvality popisuje Jingfeng Xia jako kooperaci několika procesů, mezi které patří:

- a) stanovení standardů,
- b) stanovení očekávaných, identifikovaných a odstraněných problémů,
- c) postup kontroly kvality,
- d) způsob využití geodat³⁵.

Tento dokument popisuje především způsoby měření kvality crowdsourcingových dat s cílem jejich optimalizace. Díky těmto splněným předpokladům může dojít k posílení informačních systémů na podporu vědeckých výzkumů. Autor popisuje koncept a následně také jednotlivé aspekty geodat, které by měly splňovat nástroje pro hodnocení kvality (viz obrázek 3.1)³⁵.

³⁴ RICE, Matthew T a kol. *Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data* [online]. 1. George Mason University, 2012, 114-116s. [cit. 2015-06-05]. Dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf>

³⁵ XIA, Jingfeng. *Metrics to Measure Open Geospatial Data Quality. Issues in Science and Technology Librarianship* [online]. Indiana University, 2012, 8s. [cit. 2016-03-16]. DOI: 10.5062/F4B85627. Dostupné z: <http://www.istl.org/12-winter/article1.html>



Obrázek 3.1 Konceptuální model pro hodnocení prostorových geodat získaných crowdsourcigem.
autor: Jingfeng Xia, 2012

Na Loughborough University byl publikován článek „Using VGI to enhance user judgements of quality and authority“. Autorem je Christopher J. Parker a kol. Tato studie ukázala, že mash-up aplikace využívající VGI mají zvyšující dopad na využitelnost GIS systémů. Dle autora se jedná o výrazně rychlejší a z hlediska nákladů o vysoce efektivní způsob využívání GIS. Tato práce naznačuje, že z pohledu koncových uživatelů je velmi cenný výzkum a rozvoj v oblasti VGI. To by u mash-up aplikací mohlo vést ke zvýšení jejich používání na základě projevené spokojenosti a důvěry uživatelů, kteří by využívali poskytnuté informace. V praktické části výzkum popisuje výsledky studie ve vztahu k uživatelům s podobnými potřebami při cestování. Může se jednat o potřeby pro cestování vozíčkářů, popřípadě rodičů s kočárky. Průzkum byl prováděn pro stejné cesty v Londýně³⁶.

M. van Exel publikoval článek „The impact of crowdsourcing on spatial data quality indicators“. Hlavní myšlenkou bylo vyjádřit rozdíl mezi běžnými daty, které jsou brány jako homogenní, a „volunteered geographic collections“ dat, které jsou charakterizovány jako heterogenní. Výzkum je zaměřen na tři základní ukazatele, kterými jsou prostorová činnost davu (počet editací v místě), dále časová (počet editací za časové období) a relativní (počet editací v poměru k sousedním oblastem) činnost davu. Kvantitativní výzkumy na tyto ukazatele mohou být prováděny pouze pomocí historických informací z crowdsourcových dat, které obsahují časové informace o přispívajících uživatelích. Například OSM databáze tyto informace obsahuje, takže se budou moci použít pro validaci ukazatele³⁷.

Michael McCullagh a Mike Jackson z University of Nottingham v roce 2013 publikovali článek „Crowdsourced mapping – letting amateurs into the temple?“. V článku je uvedeno, že organizace využívající aktivity přispěvatelů se rozhodly prosazovat způsob rozumného formátu školení, který je dostupný na jejich internetových stránkách. Tyto organizace prokazovaly obecně dobré výsledky ve

36 PARKER, C.J., MAY, A. and MITCHELL, V., 2012. Using VGI to enhance user judgements of quality and authority. IN: Geographical Information Science Research UK (GISRUK) 2012, 9s., Conference Proceedings.

37 van EXEL, M., E. DIAS a S. FRUIJTIER., 2012, The impact of crowdsourcing on spatial data quality indicators [online]. [cit. 2015-09-15], 4s., Dostupné z: http://www.researchgate.net/profile/Eduardo_Dias8/publication/267398729_The_impact_of_crowdsourcing_on_spatial_data_quality_indicators/links/546f49af0cf216f8cfa9d247.pdf

shromažďování údajů. Přehledné množství funkcí a zároveň dostatek příspěvatelů umožňuje provádět editaci, tudíž je zde i tendence odstranit jakékoli problémy a chyby co nejrychleji po jejich objevení uživatelem. Valná většina těchto organizací má v úmyslu používat VGI znovu³⁸.

Wikipedie a stejně tak i OSM se vyvinuly určitou formou správy; stále si však udržují pružnost a mohou se přizpůsobit měnícím se přáním VGI obyvatelstva (především co se týče nových dat, jejich struktury apod.). Právě tohoto přístupu by chtěly dosáhnout mapovací agentury, aniž by se vzdaly zjevné kontroly všech aspektů procesu mapování. Není však pochyb, že VGI reaguje rychle a obvykle pozoruhodně efektivně³⁸.

Příspěvatelé VGI mají zvláště důležitou roli při aktualizacích krizových situací, například, když dobrovolníci nacházející se v určité postižené lokalitě (například při povodních, požárech, zemětřeseních a dalších katastrofách) získávají a poskytují spolehlivé informace. Jedná se o mapování, které je rychlejší než jakákoliv jiná metoda. Je zapotřebí samozřejmě i dalších informací, aby proces fungoval správně. Může se jednat o snímky, referenční datové sady, ale také metodiky, popřípadě další dobře vzdělané dobrovolníky.

Na „7th IGRSM International Remote Sensing & GIS Conference and Exhibition“ publikoval v roce 2014 George Cho z University of Canberra článek „Some legal concerns with the use of crowd-sources Geospatial Information“. Mezi potenciální rizika při vytváření dat s příspěvím davu řadí zvýšenou problematiku týkající se kvality dat narozdíl od dat vytvářených profesionálními mapovými agenturami. Například tradičním vládním zdrojům geografických informací (GI) byla vždy přikládána větší důvěryhodnost. Autor zvažuje i možnou diskuzi nad tím, zda profesionálně školení vládní úředníci poskytují data s vyšší kvalitou než crowdsourcová společnost³⁹.

Nesporné však je, že tvorba autoritativních geografických dat je nákladná, časově náročná a prováděná obvykle ve veřejném sektoru. Na základě těchto předpokladů se zdá, že se zde zvětšuje prostor a roste potenciál pro VGI data. Zatímco na jedné straně vykazuje potenciál VGI dat vysokou přesnost, kompletnost, aktuálnost a prakticky využitelné měřítko, na straně druhé jsou tu do značné míry neznámá, nejistá a málo prozkoumaná potenciální rizika. Při využití GI je jejich „životní cyklus“ od sběru dat přes čištění až po praktické využití ohrožován chybami a nepřesnostmi. Tragické výsledky by mohly vzniknout při použití „špatných, resp. nekorektních“ dat³⁹.

Podle autora je nejvhodnějším způsobem kombinace dat ze státních organizací a dat od jejich občanů. Tento způsob byl ověřen v Kanadě. Tam jsou na obecní, respektive krajské úrovni data kombinována s daty od občanů zde žijících. S tímto způsobem experimentují také v OSM, v USGS a dalších institucích³⁹.

V roce 2014 v rámci sympózia „International Symposium of the Digital Earth – ISDE8“ publikovali N. H. Idris, M. J. Jackson a M. H. I. Ishak článek „A conceptual model of the automated credibility assessment of the volunteered geographic information“. Autoři zde vyjadřují myšlenku nutnosti využití takového nástroje, který by posoudil věrohodnost, respektive důvěryhodnost, dále správnost a soulad údajů, přičemž informace jsou

³⁸ McCULLAGH, Michael a Mike JACKSON. CROWDSOURCED MAPPING – LETTING AMATEURS INTO THE TEMPLE? [online]. 2013, 17-19s. [cit. 2015-09-15]. Dostupné z: <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-1-W1/399/2013/isprsarchives-XL-1-W1-399-2013.html>

³⁹ CHO, George. Some legal concerns with the use of crowd-sourced Geospatial Information [online]. 2014, 16s. [cit. 2015-09-15]. Dostupné z: <http://iopscience.iop.org/article/10.1088/17551315/20/1/0120401/jsessionid=DFE27F6DCDDADA>

prezentovány prostřednictvím „*Volunteered Geographic Information – VGI*“. První část studie je zaměřena na teoretický popis způsobu hodnocení kvality prostorových dat při použití konceptuálního modelu – především popisují způsob vytvoření automatizovaných procesů hledání a porovnávání dat (informací) pro jejich přesnost a správnost validace. V další části popisují samotný automatizovaný nástroj a vyhodnocují testování navrhovaného konceptuálního modelu. V rámci finálního hodnocení navrhuje autoři použít model pro posuzování dat, u nějž jsou metadata v neformálním a nestrukturovaném formátu, nebo zdroje dat nejsou k dispozici⁴⁰.

V březnu roku 2015 vydali Akash Das Sarma, Jennifer Widom ze Stanfordské univerzity a Aditya G. Parameswara z University of Illinois článek „*Globally Optimal Crowdsourcing Quality Management*“. Snaha autorů je vytvořit algoritmy pro hodnocení kvality prostorových dat. Používají několik klíčových intuitivních nápadů. Nejprve se snaží nalézt globální maximální přesnosti mapování. Na základě tohoto předpokladu dochází k vytvoření modelové matice, která je odvozena automaticky z odpovídajících maximálních přesností mapování. Dále vyvíjí nový přístup, tzv. *search*, který minimalizuje možnost výskytu potenciální chyby, respektive nepřesnosti při mapování. Na základě tohoto přístupu existuje předpoklad, že je zaručena maximální přesnost mapování. K doložení pravdivosti jimi zmiňovaných tvrzení byly provedeny experimenty na reálných a modelových (optimalizovaných) datech. V ukázkovém případě se vyskytovaly výsledky s vysokou přesností geodat. Finální část obsahuje zobecněné algoritmy pro využití v řadě případových studií, které využívají různé formy pravděpodobnostních matic přesnosti. Autoři poukazují na to, že je důležité si uvědomit, že jde o zobecňující formu rozšíření algoritmů. Ne všichni uživatelé však poskytují odpovědi se stejnou, vysokou kvalitou přesnosti. V tomto případě mohou být algoritmy pro práci v GIS neadekvátní. Na základě předpokladů uvedených a ověřených v tomto článku je možné vytvářet další efektivní algoritmy pro hodnocení kvality prostorových dat získaných crowdsourcingovým mapováním⁴¹.

Obecným hodnocením kvality prostorových dat pro zvolené typy prostorových analýz se v diplomové práci zabývá Petr Šrámek (2011), kterou vypracoval na univerzitě v Pardubicích. Cílem této práce bylo definovat kritéria pro určení kvality dat a vytipování příkladů, které využívají prostorové analýzy. V této práci autor obecně definuje pojem prostorová data, dále popisuje jejich složky a porovnává digitální a analogový formát prostorových dat. Byly popsány jak kvantitativní, tak kvalitativní parametry, které vycházejí z norem kvality prostorových dat. Jedna z případových studií zaměřenou na kvalitu dat byla zacílena na zjišťování vhodné lokality pro pěstování vinné révy v okolí Karlštejna. Přístup k informacím o kvalitě prostorových dat přitom autor shledal poměrně nedostatečným. Vyjadřuje nespokojenost s obtížnou, popřípadě žádnou možností dohledatelnosti informací o parametrech kvality. Důkazem tohoto tvrzení je fakt, že při pořizování prostorových dat jsou mnohdy uživatelům dostupná pouze stručná a nedostatečně definovaná metadata⁴².

⁴⁰ IDRIS, N. H., M.J. JACKSON a M.H.I. ISHAK. A conceptual model of the automated credibility assessment of the volunteered geographic information [online]. 2014, 4-6s., [cit. 2015-09-15]. Dostupné z: <http://iopscience.iop.org/article/10.1088/1755-1315/18/1/012070/meta>

⁴¹ DAR SARMA, Akash, Aditya G. PARAMESWARAN a Jennifer WIDOM. *Globally Optimal Crowdsourcing Quality Management* [online]. 2015 [cit. 2015-09-15]. Dostupné z: <http://ilpubs.stanford.edu:8090/1111/>

⁴² ŠRÁMEK, Petr. *Hodnocení kvality prostorových dat pro zvolené typy prostorových analýz*. Pardubice, 2011. Diplomová práce. Univerzita Pardubice. 17-40s. Vedoucí práce Doc. Ing. Jitka Komárková, Ph.D.



Obrázek 3.2 Grafické vyjádření vhodných lokalit pro pěstování vinné révy v okolí Karlštejna.
autor: Bc. Petr Šrámek, 2011

Problematicke kvality prostorových dat se v jedné z částí disertační práce „*Vizualizace nejistoty v environmentálních studiích*“ (2013) věnuje také Jan Brus. Uvádí, že problematické už je samo vymezení – doplnit „vymezení ČEHO?“, protože často dochází k problémům s názvoslovím a samotným slovním spojením. V další části řeší autor dimenze prostorových dat a následně jejich kvalitu. Vymezuje pojem „kvalita“ a také zde uvádí veškeré důležité standardy. Tyto standardy jsou všeobecně využívány pro stanovení kvality prostorových dat. V poslední části této kapitoly jsou vysvětlovány a detailně popsány jednotlivé komponenty kvality prostorových dat. Je zde uvedeno, že kvalita prostorových dat je multikomponentní. V tomto případě pojem vyjadřuje předpoklad, že neexistuje pouze jedno pojetí prvků kvality prostorových dat. Tato skutečnost je demonstrována relativně vysokým počtem standardů pro hodnocení kvality prostorových dat⁴³.

Další publikace věnující se moderním způsobů mapování a mimo jiné i metodě crowdsourcingu je „*GeoParticipace – Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*“, kterou publikoval v roce 2014 Jiří Pánek a kol. Cílem knihy není pouze přesně definovat, co GeoParticipace je, ale především nastinit široké možnosti využití geoinformačních technologií v této oblasti. Pro účely této diplomové práce je důležitá kapitola „*Online boj s nelegálními skládkami (M. Kubásek a J. Pánek)*“⁴⁴.

V této části je nejprve vysvětlen teoretický základ metody crowdsourcingu a její úskalí. Následně jsou popsány další aplikace podporující tuto metodu. Finální část popisuje aplikaci „ZmapujTo“, jejímž autorem je Miroslav Kubásek. Tato aplikace je kompletně založená na metodě crowdsourcingu. Autor v publikaci vysvětluje samotnou

⁴³ BRUS, Jan. *Vizualizace nejistoty v environmentálních studiích*. Olomouc, 2013. disertační práce 22-31s. UNIVERZITA PALACKÉHO V OLOMOUCI. Přírodovědecká fakulta

⁴⁴ PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 27-39s. 978-80-244-4359-1

podstatu projektu, jejímž cílem je bojovat proti nelegálním skládkám odpadu v České republice. Komplexně je projekt zaměřen na občany, kteří nejsou spokojeni s výskytem černé skládky v jejich okolí a chtějí s tím něco udělat pomocí jednoduché a tzv. „UserFriendly“ aplikace, do které nahlásí přesnou polohu skládky (doplněnou o fotografii skládky). Následně je toto oznámení odesláno příslušnému obecnímu, respektive městskému úřadu dané obce (obrázek 3.3)⁴⁵.



Obrázek 3.3 GeoParticipace.
autor: Mgr. Jiří Pánek, Ph.D, 2014

3.4 Využití crowdsourcingu v praxi

V současnosti je již zcela běžnou součástí lidské společnosti využívání mobilních aplikací (existují však také verze aplikací běžně dostupných na PC). Většina aplikací je využívána právě pro svůj, zpravidla jednoduchý, účel. Již málo uživatelů uvažuje nad aplikací jako o prostředku pro rychlý, levný a efektivní sběr prostorových dat pomocí metody crowdsourcingu. V tabulce 3.2 jsou aplikace děleny podle způsobu jejich využití.

Tabulka č. 3.2 Příklady aplikací využívají princip metody⁴⁶

Úkol / účel	Popis	Příklad
Zobrazování	Vytváření kolekcí obrázků.	<ul style="list-style-type: none"> Grassroots Mapping
Georeferencování	Rektifikace snímků a map.	<ul style="list-style-type: none"> Grassroots Mapping NYPL Map Rectifier
Transkripce (přepis)	Převod zdrojových textů do digitální podoby.	<ul style="list-style-type: none"> OldWeather
Digitalizace	Sběr geoprostorových a atributových dat z mapy nebo snímku.	<ul style="list-style-type: none"> OSM Google MapMaker Wikipedie
Vkládání atributů	Přidávání popisné informace (atributů) do datové sady.	<ul style="list-style-type: none"> Galaxy Zoo
-	-	-

⁴⁵ PÁNEK, Jiří a kol. GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 27-39s. 978-80-244-4359-1

⁴⁶ RICE, Matthew T a kol. Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data [online]. 1. George Mason University, 2012, 32-33s.[cit. 2015-06-05]. Dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf>

<i>Hlášení</i>	Shromažďování informací o určitém místě pomocí pozorování, popřípadě pomocí mobilních zařízení.	<ul style="list-style-type: none"> • Louisiana Bucket Brigade • GasBuddy • Street Bump • Syria Tracker
<i>Vyhledávání</i>	Identifikace konkrétních tvarů v mapě, popřípadě na snímku.	<ul style="list-style-type: none"> • DARPA Red Ballon • Field Expedition Mongolie
<i>Sledování</i>	Zaznamenávání trasy, obvykle pomocí GPS.	<ul style="list-style-type: none"> • Waze
<i>Ověřování</i>	Ověřování kvality stávající prostorové informace.	<ul style="list-style-type: none"> • NAVTEQ Map Reporter • Geo-Wiki.org • OSM Inspector
<i>Dotazování / Zeměměřičství</i>	Sběr geodat na základě informací od uživatelů.	<ul style="list-style-type: none"> • SurveyMapper
<i>„Sociální“ mapování</i>	Získání geoprostorové informace z příspěvků na sociálních sítích.	<ul style="list-style-type: none"> • Twitter • LinkedIn
<i>Sdílení</i>	Umístění obsahu na hostované stránky, potenciálně včetně dat, aplikací, popřípadě mapy, kde mohou mít uživatelé přístup.	<ul style="list-style-type: none"> • ArcGIS Online • GeoCommons

V zahraničí se aplikace pro vytváření prostorových dat pomocí crowdsourcingu v současnosti používají již naprosto běžně. V zahraničí je poměr chytrých telefonů (s operačními systémy iOS, Android a Windows Phone) ke klasickým mobilním telefonům markantně vyšší, než je tomu v případě České republiky. Je však potřeba brát ohled o jakou část světa se jedná (existují velké rozdíly mezi Evropou, Severní a Jižní Amerikou a Asií).

OpenStreetMap – OSM

Ve společnosti zabývající se prostorovými daty je průkopnickou crowdsourcingovou aplikací „*OpenStreetMap*“. Projekt datuje svůj vznik v roce 2004 a přišel s ním britský absolvent informatiky Steve Coast. Velké změny přinesl rok 2007, v němž počet uživatelů narostl na 20 000. V současnosti má projekt více než jeden milion přispěvatelů. OSM je projekt tvořený komunitou uživatelů, kteří přidávají a udržují data o silnicích, cestách, restauracích, obchodních centrech a jiných dalších lokalitách po celém světě. Vychází z několika základních předpokladů⁴⁷:

- Místní znalost: OSM klade důraz na místní znalost. Přispěvatelé využívají jak letecké snímky, GPS přístroje, tak klasické mapy, aby ověřili, že je aplikace OSM přesná a aktuální.
- Řízeno komunitou: Komunita je různorodá a každý den se rozrůstá. Mezi přispívatele patří odborná kartografická či

⁴⁷ RICE, Matthew T a kol. *Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data* [online]. 1. George Mason University, 2012, 39s. [cit. 2015-06-05]. Dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf>

geoinformatická veřejnost, technici spravující servery OSMA humanitní pracovníci mapující oblasti zasažené katastrofou. Značně rozsáhlou část přispěvatelů tvoří amatérští přispěvatelé.

- c) Otevřená data: OSM jsou *svobodná* data, která jsou šířena za podmínek „*Open Data Commons Open Database License (ODbL)*“. Data lze kopírovat, upravovat a distribuovat, pokud je uveden jako zdroj OSM a jeho přispěvatelé. Pokud jsou data použita v jiném díle, je potřeba toto dílo šířit pod stejnou licenci jako OSM. Kartografická díla v OSM mapových dlaždicích a dokumentace jsou šířeny pod licenci „*Creative Commons — Attribution-ShareAlike 2.0 (CC-BY-SA)*“.
- d) Právní informace: Web, mobilní aplikace a mnoho souvisejících služeb za komunitu oficiálně provozuje „*OpenStreetMap Foundation*“ (OSMF)⁴⁸.

Od počátku projektu jsou data pořizována dobrovolníky, kteří systematicky mapují pomocí ručních GPS přijímačů. GPS data jsou zpracována na počítači a následně nahrána do geodatabáze OSM. Mapovat lze při procházkách, na kole, v autě, při cestování v zahraničí a podobně. Často jsou organizovány tzv. „*mapping parties*“, které jsou zaměřené na skupinové mapování dané oblasti, často jako součást konference⁴⁶.

Projekt svými daty podpořily i některé z velkých firem – silniční síť USA byla tak například zmapována projektem TIGER. Na území České republiky přispěl daty např. Ústav pro hospodářskou úpravu lesa (ÚHUL).

Před pěti lety, v roce 2010, vznikl tzv. „*Humanitarian OpenStreetMap Team*“ – zkráceně HOT. Když v roce 2010 došlo k zemětřesení na Haiti, následovalo rychlé nasazení velkého množství dobrovolníků při tvorbě map. HOT se dále zapojilo do různorodých druhů humanitárních akcí, například po tajfunu Haiyan nebo při vypuknutí epidemie eboly v západní Africe. Jednou z posledních akcí bylo využití HOT při zemětřesení v Nepálu 2015. V případě, že nelze mapovat přímo v terénu, využívá se obkreslování ze snímků⁴⁹.

Syria Tracker

Za vznikem aplikace „*Syria Tracker*“ stojí skupina araboameričanů žijících v USA. Aplikace byla spuštěna v dubnu roku 2011, což bylo pár týdnů po propuknutí prvních nepokojů v Sýrii. Samotná aplikace je vytvořena na základě kombinace funkcionality cloudové aplikace „*Crowdmap od Ushahidi*“ spolu s upravenou verzí „*HealthMap*“, tzv. HealthMap Crisis. Právě kombinace těchto dvou technologií umožňuje plně využít princip crowdsourcingu (reporty událostí pomocí sms, webu, e-mailu, aplikace, či Twitteru). Potenciál zde má i data-mining, který je používán k monitoringu domácích i zahraničních médií⁵⁰.

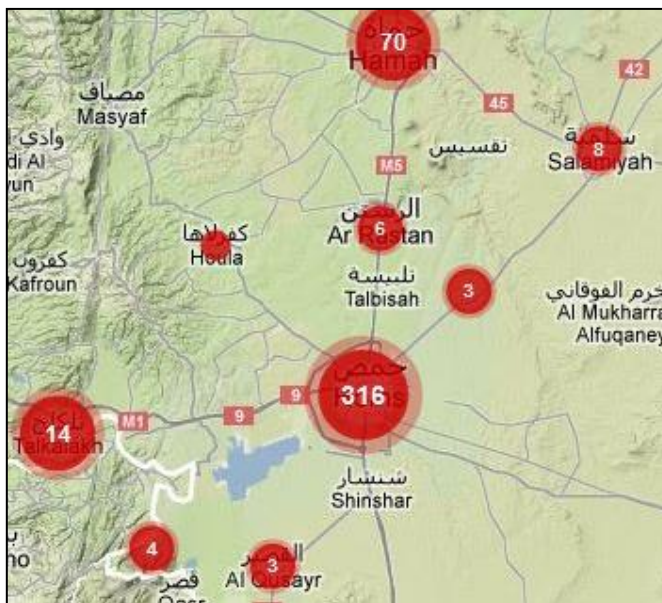
⁴⁸ OpenStreetMap:OSM [online]. 2014 [cit. 2015-06-07]. Dostupné z: <https://www.openstreetmap.org/about>

⁴⁹ Humanitarian Open Street Map Team: Ongoing projects [online]. 2015 [cit. 2016-03-17]. Dostupné z: <https://hotosm.org/projects/ongoing>

⁵⁰ *Syria Tracker: HUMANITARIAN TRACKER* [online]. 2011 [cit. 2015-06-07]. Dostupné z: <http://www.humanitariantracker.org/#!/syria-tracker/cj00>



Obr. 3.4 Syria Tracker.



Obr. 3.5 Ukázka zobrazení příspěvků v mapě.

zdroj: <http://www.gisportal.cz/2012/08/syria-tracker-crowdsourcing-v-akci/>

Tomnod

Tomnod je projekt, který provozuje satelitní společnost „DigitalGlobe“ využívající crowdsourcing k identifikaci objektů a míst na satelitních snímcích. Svě využití tato aplikace nachází v mnoha směrech. Například, když v roce 2011 Tomnod spolupracoval na detekci a následném počítání přístřešků pro uprchlíky. Uživatelé byli požádáni, aby za pomoc satelitních snímků poskytnutých DigitalGlobe počítali přístřešky pro uprchlíky. Dalším počinem bylo mapování území, které bylo poškozeno cyklónem Haiyan.

Jedním z největších projektů, kterých se účastnila aplikace Tomnod bylo vyhledávání trosk letadla Boeing 777-200 společnosti Malaysia Airlines. Letadlo havarovalo v březnu roku 2014. Společnost DigitalGlobe měla k dispozici nejnovější snímky oblastí s nejpravděpodobnějším místem pádu letadla. Úkolem uživatelů bylo si poskytnutý snímek prohlédnout a v případě identifikace podezřelého předmětu jej zaznamenat pomocí vybraných kategorií (olejová skvrna, záchranný člun, trosky letadla, popřípadě ostatní). Pokud je některé z míst označeno nezávisle několika uživateli a je pomocí algoritmu CrowdRank vyhodnoceno jako podezřelé, pak se na řadu dostávají profesionální analytici, kteří ho prozkoumají důkladněji⁵¹.

Waze

Původně izraelský projekt „Waze“ je automobilová navigace, jejíž obsah vytvářejí sami uživatelé. V současnosti je vlastníkem společnosti americký Google. Jedná se o webovou službu a zároveň mobilní aplikaci. Uživatelé samotné aplikace tvoří veškerý obsah navigace včetně mapových podkladů, což je zároveň největší slabinou celého konceptu. Tato nevýhoda se mění na jeden z největších kladů celého projektu ve chvíli, kdy Waze začne využívat velkého množství lidí. Ulice jsou vytvářeny po prvních průjezdech se zapnutou aplikací a po následných úpravách a korekcích ve webovém

⁵¹ Pomozte najít malajsijský boeing [online]. Olomouc: GISportal.cz, 2014 [cit. 2016-03-17]. Dostupné z: <http://www.gisportal.cz/2014/03/pomozte-najit-malajsijsky-boeing/>

rozhraní. Uživatelé však nemapují pouze uliční síť, ale informují další uživatele o aktuálních kolonách, policejních radarech a jiných dopravních omezeních⁵².

Ushahidi

Ushahidi (ve svahilštině „svědectví“) funguje jinak než například Open Street Map nebo Waze. Soustředí se na mimořádné události, jako jsou povodně, zemětřesení a nepokoje. Zaměřuje se na přesně definovanou oblast po kratší dobu. Cílem je koncentrovat aktuální informace, fotografie a videa vztahující se k události v mapě. Obsah je vytvářen jak samotnými uživateli, tak automatickým načítáním ze služeb jako Flickr, YouTube nebo Twitter. Tento přístup využívá geotagování příspěvků. K praktickému nasazení došlo při zemětřesení v Japonsku a na Novém Zélandu 2011 či při nepokojích v Libyi⁵³.

V České republice se mobilní aplikace pro sběr geoprostorových dat začaly intenzivněji využívat s příchodem „*smartphonů*“ či „*chytrých telefonů*“, obecně řečeno - s příchodem mobilních zařízení. Mezi mobilní zařízení patří právě výše uvedené mobilní telefony, smartphony, tablety a další. Podle statistik operátora T-Mobile podíl chytrých telefonů v ČR naroste ze současných 59 % přibližně k 79 %. Je zajímavé, že 20 % uživatelů chytrých telefonů vlastní současně i tablet. V další části kapitoly jsou uvedeny vybrané aplikace využívající metodu crowdsourcingu pro sběr geodat.

Dej Tip

Mobilní aplikaci „*Dej Tip*“ pro hlášení závad vytvořila společnost Gefos ve spolupráci s firmou Intergraph CS. Projekt běží od dubna roku 2013. Aplikace Dej Tip je založená na kooperaci chytrých telefonů a internetových služeb. DejTip je kompatibilní se systémy iOS (Apple) a Android (Google). Poprvé byla aplikace nabídnuta zdarma v Českých Budějovicích, ale v současné době je již plně rozšířena ve velkém množství měst po celém území České republiky, jako například v Havířově, Ostravě, Pardubicích a dalších. Občané mají prostřednictvím této aplikace možnost upozornit na problémy v daném městě. Součástí hlášení o závadě je fotografie, dále lokační údaje z mobilní polohové služby, komentář dané závady, informace o použitém operačním systému a typu mobilního zařízení. Pokud není možné přesně určit polohu, může uživatel zpřesnit polohu pomocí kliknutí do mapy. Největší výhodou aplikace je, že občan může hlášení odeslat pomocí chytrého telefonu s připojením k internetu přímo z místa závady. Pomocí webového portálu může zaslátel nebo i kdokoli jiný sledovat odstranění závady. Aplikace přispívá k usnadnění a urychlení vyřešení závady (obrázek 3.6)⁵⁴.

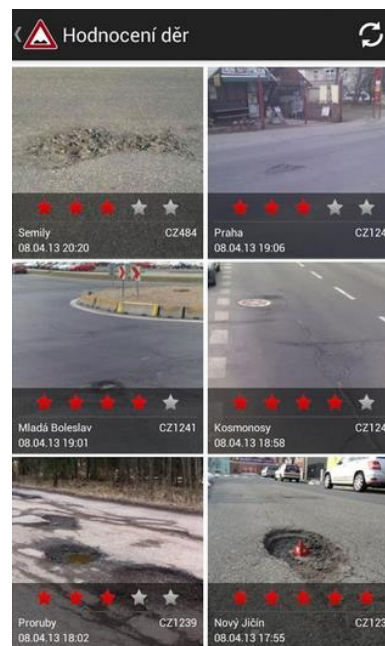
⁵² Crowdsourcing: síla davu: Waze [online]. Olomouc: GISportal.cz, 2011 [cit. 2016-03-17]. Dostupné z: www.gisportal.cz/2011/03/crowdsourcing-sila-davu/

⁵³ Crowdsourcing: síla davu: Ushahidi [online]. Olomouc: GISportal.cz, 2011 [cit. 2016-03-17]. Dostupné z: www.gisportal.cz/2011/03/crowdsourcing-sila-davu/

⁵⁴ HINTNAUS, Ivo. GISPORTAL.CZ. DejTip, Výmoly, Citysources: crowdsourcingové geoaplikace [online]. 2013 [cit. 2015-06-07]. Dostupné z: <http://www.gisportal.cz/2013/11/dejtip-vymoly-citysources-crowdsourcing-geoaplikace/>



Obr. 3.6 prostředí aplikace DeJTIP.



Obr. 3.7 prostředí aplikace Výmoly.

zdroj: <http://www.gisportal.cz/2013/11/dejtip-vymoly-citysourced-crowdsourcingove-geoaplikace/>

Výmoly

Projekt „Výmoly“ je zaměřen na monitorování kvality silnic. Autoři označují Výmoly jako unikátní motoristický projekt mapující stav českých silnic. V roce 2013 se projekt rozšířil také na území Slovenska a Maďarska. V České republice s tímto projektem spolupracuje například pojišťovna Generali nebo Český rozhlas Radiožurnál v rámci dopravního zpravodajství „Zelená vlna.“. Upozornění na výmol mohou lidé přidávat nejen pomocí webového portálu, ale také pomocí aplikace v mobilním telefonu. Aplikace je kompatibilní s operačními systémy iOS (Apple) a Android (Google). Projekt Výmoly pomáhá zlepšovat aktuální stav silnic a zajišťuje aktivní komunikaci mezi řidiči a uživateli. V loňském roce identifikoval přibližně 3500 silničních výtluků. Pozitivním zjištěním je, že 40 % bylo opraveno. Přes počáteční nedůvěru dnes silničáři přiznávají, že projekt Výmoly je vítaným pomocníkem a zdrojem prostorových dat (obrázek 3.7)⁵⁵.

BioLog

Jedním z netradičních způsobů získávání prostorových dat crowdsourcingem je projekt „BioLog“. Autorem projektu je Agentura ochrany přírody a krajiny ČR (AOPK ČR), jmenovitě Jan Zárybnický. Jedná se o jednu z možností sběru informací o výskytu druhů živočichů, rostlin, hub a dalších organismů především v České republice. Aplikace BioLog byla vytvořena pro zájemce o přírodu a slouží jako digitální zápisník fauny a flóry. Umožňuje lokalizovat polohu (aplikace nevyžaduje připojení k internetu, stačí GPS) a popsat pozorování druhů přímo v terénu, přidat fotografii a uložit vše do mobilního zařízení. Pozorování si lze zobrazit v mapě. Vybraná pozorování je možné odeslat do databáze výskytu druhů, která je spravována AOPK ČR. Data lze také exportovat (ve formátu CSV) pro vlastní využití a vyhodnocení⁵⁶.

⁵⁵ HINTNAUS, Ivo. GISPORTAL.CZ. DeJTIP, Výmoly, Citysources: crowdsourcingové geoaplikace [online]. 2013 [cit. 2015-06-07]. Dostupné z: <http://www.gisportal.cz/2013/11/dejtip-vymoly-citysourced-crowdsourcingove-geoaplikace/>

⁵⁶ AGENTURA OCHRANY PŘÍRODY A KRAJINY ČR, Jan Zárybnický. BioLog [online]. Praha, 2015 [cit. 2015-06-07]. Dostupné z: <http://biolog.nature.cz/cz/Article/AboutApp#1>

Plznito

Aplikace s názvem „Plznito“ nabízí lidem novou možnost, jak upozornit na poničené dětské houpačky, chybějící značky, nefunkční osvětlení, popřípadě na další závady ve veřejném prostoru města. Aplikace umožňuje zadat nové hlášení a poté informuje o vyřešení daných podnětů. Koncem února 2016 byl stav hlášení: 65 v řešení, 890 vyřešených a 45 nepatří městu. Aplikace je další možným způsobem komunikace mezi občanem a městem. Lidé mohou požadavky zadávat pomocí mobilního telefonu, do něhož si zdarma stáhnou aplikaci „Plzeň-občan“ (je dostupná pro operační systémy iOS, Android), nebo prostřednictvím webových stránek plznito.cz nebo plzni.to⁵⁷.

3.4.1 ZmapujTo

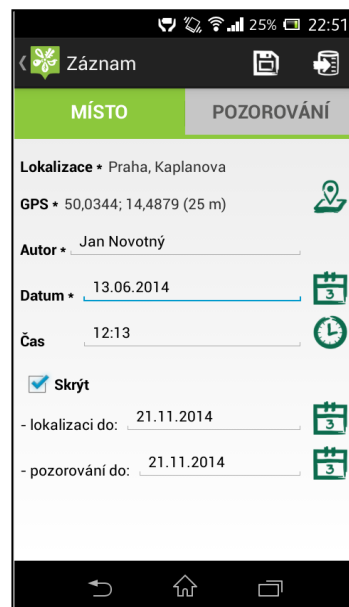
Autorem projektu „ZmapujTo“ je český environmentalista a vysokoškolský pedagog z Masarykovy univerzity v Brně Miroslav Kubásek. První verze vznikla v květnu 2012 jako ekologický projekt s primárním cílem bojovat proti nelegálním skládkám odpadu v České republice. Projekt je zaměřen na občany, kteří jsou nespokojeni s výskytem černých skládek v jejich okolí. Již první verze byla uživatelsky velmi jednoduchá. V březnu roku 2014 byla spuštěna druhá verze této aplikace. Primární cíl projektu, tedy boj proti nelegálním skládkám, zůstal stejný. Rozšíření aplikace spočívá především v možnosti hlášení dalších problémů, s nimiž se můžeme ve městech a v přírodě setkat. Zasláné podněty jsou automaticky přiřazovány a zasilány místním samosprávám (uživatel však zůstává v anonymitě). Díky tomuto způsobu řešení může uživatel sledovat vývoj jednotlivých kauz, případně se aktivně zapojit do jejich řešení. Aplikace je kompatibilní s operačními systémy iOS (Apple), Android (Google) a s nástupem druhé verze také s Windows Phone (Microsoft). S tímto projektem spolupracuje více než 900 subjektů, mimo jiné města Olomouc, Praha, Zlín, nebo Brněnské komunikace a.s.⁵⁸.

Problematika nelegálních skládek

Černou skládkou se rozumí odkládání odpadů na místa, která k tomu nejsou určena. Zpravidla se jedná o místa, která jsou dobře dostupná a jsou v místech obtížné vizuální kontroly (je zde malé riziko, že někdo bude při vyhazování odpadu zahlédnut). Za černou skládku je také považováno umístování odpadů k jiným druhům kontejnerů, než pro které jsou určeny.

Legislativa

Problematika černých skládek je v českém odpadovém hospodářství diskutována již mnoho let. Zásadní důvodem pro diskusi je právě problematika, v řadě případů téměř nemožná identifikace subjektu, který takovou skládku založil, tedy subjektu, u něhož



Obr. 3.8 Aplikace BioLog.
zdroj: <http://biolog.nature.cz/>

⁵⁷ Plznito [online]. Plzeň: Správa informačních technologií města Plzně, 2016 [cit. 2016-03-17]. Dostupné z: www.plznito.cz; www.plzni.to

⁵⁸ PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 27-39s. 978-80-244-4359-1

by bylo možné vyvodit odpovědnost za vznik takovéto skládky. V případě jednodušší identifikace by bylo možno na základě této odpovědnosti uložit povinnost takovou nelegální skládku odstranit, případě za založení nelegální skládky uložit sankci.

Současná právní úprava, to znamená *zákon č. 185/2001 Sb., o odpadech a o změně některých dalších zákonů*, ve znění pozdějších předpisů, tuto problematiku komplexně neupravuje. Vzhledem k tomuto faktu byla pro odstraňování černých skládek více efektivní právní úprava předchozí, tj. *zákon č. 125/1997 Sb., o odpadech*. Tento zákon, na rozdíl od současně platného a účinného zákona, řešil otázku nelegálních skládek výslovně tak, že odpovědnost za nelegálně shromážděné odpady na určitém pozemku spadala na jeho vlastníka.

Zakládání černé skládky je přestupkem. Fyzické osobě hrozí pokuta od 3 000 Kč do 50 000 Kč. Firmám, respektive právnickým osobám hrozí postih až do výše 10 milionů korun. Postih za nedovolené nakládání s odpady se liší také podle původců odpadů.

Nadějí je v této problematice nově chystaný odpadový zákon, který se vrací k původní myšlence *zákona č. 125/1997 Sb., o odpadech*. Zákon umožní získat obcím s rozšířenou působností více pravomocí při likvidaci černých skládek a majitelům pozemků uloží povinnost řádně se o svůj majetek starat.

Současná verze ZmapujTo

V březnu 2014 byla spuštěna druhá verze ZmapujTo.cz. Cíl projektu bojovat proti nelegálním skládkám odpadu a zapojit občany zůstal stejný. Oproti prvotní verzi je nejdůležitější změnou možnost hlásit nejen černé skládky, ale také celou řadu problémů, na které lze v přírodě či ve městě narazit (přeplněný odpadkový koš, poškozená lavička, nebezpečná stavba a další)⁵⁹.

Pro mapování bylo potřeba zvolit moderní, efektivní a široce rozšířenou platformu. Z tohoto důvodu byla vybrána mobilní aplikace a byl zpřístupněn interaktivní webový formulář pro zasílání hlášení. Celý web je koncipován tak, aby byl rychlý a intuitivní. Mobilní aplikace využívá responsivního designu, což umožňuje web www.ZmapujTo.cz používat například na tabletu či chytrém telefonu a zůstane přitom zachována plnohodnotná funkčnost. Mobilní aplikaci lze zdarma stáhnout pro všechny tři typy nejběžnějších operačních systémů (Android, iOS, Windows Phone)⁵⁹.

Použité technologie

Jednotlivé regiony jsou v aplikaci třízeny dle dat geodatabáze ArcČR 500 (verze 3.1), která je dispozici zdarma. Pro převod dat do souřadnicového systému WGS 84 byl použit software QGIS (verze 2.0 – Dufour). Pro zobrazování jednotlivých hlášení v mapě je využito „*Google Maps API*“. V aplikaci je současně implementováno „*Google Street View*“ umožňující uživatelům prohlédnout si dané místo a usnadnit si tím orientaci. Detailní náhled dává k dispozici také katastrální mapu s přímým spojením na web ČÚZK⁶⁰.

Responzivní design webu www.ZmapujTo.cz je postaven šabloně „*Twitter Bootstrap*“ a naprogramován pomocí knihovny „*AngularJS*“ jako single-page aplikace s propojením na Google Maps API. Pozadí (backend) serveru běží na „*NodeJS*“ (událostmi řízený runtime založený na jazyce JavaScript) a „*Express*“ (aplikační framework nad NodeJS

⁵⁹ PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 28s. 978-80-244-4359-1

⁶⁰ PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 30-31s. 978-80-244-4359-1

pro server). Pro uložení dat se využívá dokumentově orientovaná „NoSQL databáze Mongo“⁶¹.

Mobilní aplikace je postavena na technologiích „Sencha Touch“ a „Apache Cordova“, jejichž API je postavené nad programovacím jazykem PHP s využitím frameworku „Symphony, Doctrine, ODM“. Pro zaslání notifikačních e-mailů je používána platforma „Mandrill“⁶¹.

Server je hostován na platformě „DigitalOcean“, která nabízí využití cloudových serverů běžících na SSD discích v ceně již od 5\$ za měsíc. Tato platforma umožňuje zvládnout také například jednorázový nápor, čímž jsou minimalizovány výpadky provozu dané aplikace⁶¹.

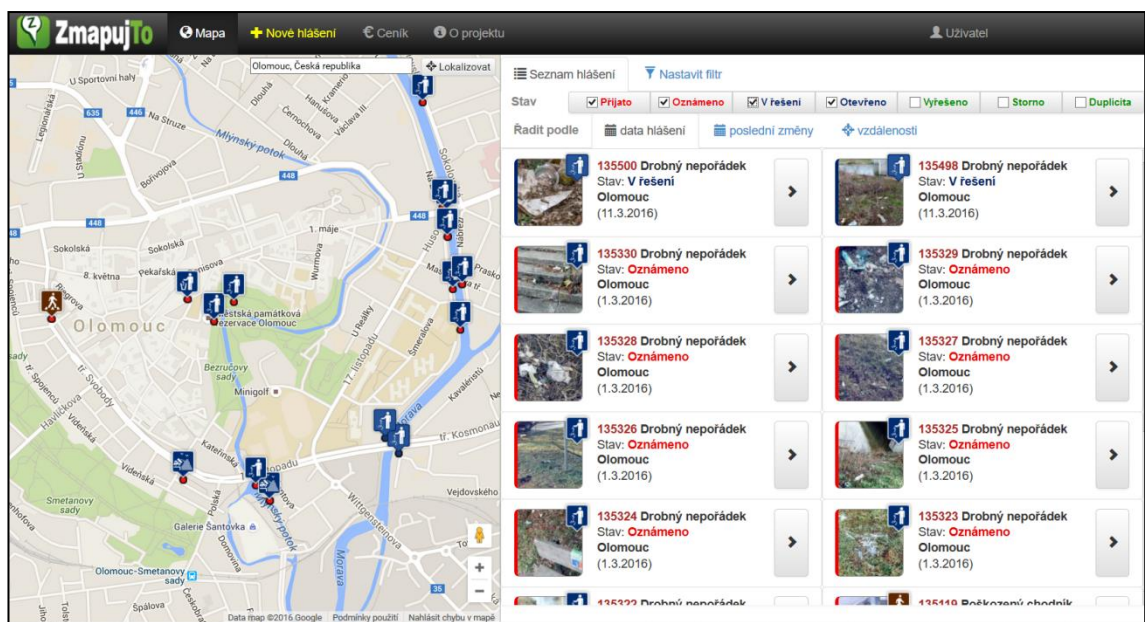
Výstupy

Za hlavní výstup projektu je považována webová stránka www.ZmapujTo.cz, která má k dispozici interaktivní mapu České republiky, v níž jsou zaznamenána jednotlivá hlášení. Interaktivní mapa je rozdělena na dvě části. V levé části se nachází hlavní mapa obsahující jednotlivá hlášení. Pravá část zobrazuje podrobný seznam jednotlivých hlášení (obrázek 3.9)⁶¹.

Zaznamenaná hlášení lze libovolně filtrovat a tím si výpis omezit pouze na ta hlášení, která uživatele zajímají. Jednou z možností je filtrovat hlášení podle aktuálního stavu, které jsou ve ZmapujTo dělena do následujících stavů:

- Přijato – hlášení bylo přijato do systému ZmapujTo.
- Oznámeno – o hlášení byla informována příslušná samospráva, popřípadě jiná organizace (nejčastěji pomocí e-mailu).
- V řešení – hlášením se začala příslušná organizace zabývat (skládka k odklizení).
- Otevřeno – hlášení se nepodařilo vyřešit (např. nedohledatelný vlastník).
- Vyřešeno – hlášení bylo úspěšně vyřešeno (např. skládka byla odklizená).
- Storno – hlášení bylo stornováno. Nejčastěji se jedná o chybné, či nesmyslné hlášení.
- Duplicita – hlášení je duplicitní. Další hlášení jsou přiřazována k jednomu „hlavnímu“ hlášení. Vzniká tak historie hlášeného problému a lze tak sledovat jeho časový vývoj⁶¹.

⁶¹ PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 30-31s. 978-80-244-4359-1



Obrázek 3.9 Grafické prostředí webu www.ZmapujTo.cz.
autor: J. Vrkoč (2016)

Další možností je výběr hlášení podle toho, o jaký typ se jedná. K dispozici je pět základních předdefinovaných skupin. Tyto skupiny jsou následně děleny na další typy hlášení (některé typy hlášení vyžadují specifikaci dalších parametrů jako velikost skládky, jaké druhy odpadu se v ní nachází a další). Současně je možné hlášení filtrovat podle „Stáří hlášení“. Jednotlivé skupiny a typy hlášení jsou uvedeny v obrázku 3.10⁶².

<p>Města, obce, veřejné prostranství</p> <ul style="list-style-type: none"> Poškození zastávky MHD, autobusů, vlaků Poškozené, nefunkční veřejné osvětlení Poškozená kanalizace Poškozená elektrická infrastruktura Nebezpečná stavba Rušení veřejného pořádku, obtěžování <p>Doprava, silnice, cyklostezky</p> <ul style="list-style-type: none"> Poškozený chodník, cesta Poškozený povrch vozovky Poškozená svodidla, mosty Překážka na silnici, u silnice Poškozená dopravní značka, zrcadlo Poškozený semafor Špatné dopravní značení na komunikaci 	<p>Odpady, čistota</p> <ul style="list-style-type: none"> Drobný nepořádek Černá skládka Odpadkový koš, kontejner Odstavený autovrak <p>Příroda, parky, hřiště, zvířata</p> <ul style="list-style-type: none"> Poškození lavičky, zábradlí Poškozené pískoviště, hřiště, sportoviště Poškozená socha, památník Závada na veřejné zeleni Zvířata, týrání <p>Jiná hlášení</p> <ul style="list-style-type: none"> Návrh na zlepšení Jiné hlášení, problém
--	---

Obrázek 3.10 Filtrování podle skupin a dle jednotlivých typů hlášení.
autor: J. Vrkoč (2016)

⁶² PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 33s. 978-80-244-4359-1

Kontrola kvality dat ve ZmapujTo

Kvalita jednotlivých reportů je v aplikaci ZmapujTo zajišťována expertně. V současnosti je každé hlášení ještě předtím tím, než se objeví v systému *ZmapujTo.cz*, manuálně zkontrolováno administrátorem. Administrátorem je autor samotné aplikace, Miroslav Kubásek. Následně je stav hlášení změněn na „Přijato“. Jedná se tedy o plně manuální administraci (kontrolu) reportů, která je časově dosti náročná. Je kontrolována jak atributová přesnost (fotografie), tak polohová přesnost (StreetView).

V rámci praktické části diplomové práce byl vytvořen distribuční model hodnotící polohovou přesnost jednotlivých hlášení velkých černých skládek v aplikaci ZmapujTo. Výstupem modelu je rastr reprezentující pravděpodobnost výskytu černých skládek na území hlavního města Prahy. Postup zpracování a výsledky jsou popsány v kapitole 4, respektive v kapitole 5. Součástí výsledků je návrh konceptu pro komplexní hodnocení kvality dat aplikovatelný na příkladu aplikace ZmapujTo.

Otevřená platforma

Aplikace ZmapujTo je primárně zaměřená na města a obce. Využívá otevřenou platformu, což znamená, že je tento nástroj možné využít pro mapování dalších dílčích projektů. Na jaře 2014 byla tato platforma využita pro monitoring billboardů před volbami do europarlamentu. Tato akce proběhla ve spolupráci s organizací Transparency International – Česká republika, o.p.s. Následně byla aplikace opět využita pro monitorování kampaní před komunálními a senátními volbami v roce 2014⁶³.

⁶³ PÁNEK, Jiří a kol. *GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?*. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 38s. 978-80-244-4359-1

4 VLASTNÍ ŘEŠENÍ

Při snaze o zajištění kvality prostorových dat je ve světě nejčastěji využíván přístup založený na kontrolních algoritmech. Součástí zadání diplomové práce je požadavek na navržení vhodného konceptu pro ověření kvality vybraných geodat. Po konzultaci s vedoucím práce bylo rozhodnuto, že bude prakticky navržen, optimalizován a testován distribuční model, který zajistí hodnocení polohové přesnosti na příkladu aplikace ZmapujTo. Kapitola 4 popisující navržený postup pro hodnocení polohové kvality hlášení lze rozdělit do několika podkapitol:

- příprava vstupních dat,
- výpočet v softwaru Maxent,
- testování a optimalizace.

4.1 Příprava vstupních dat

Samotná část přípravy vstupních dat spočívala v úpravě dat do podoby, kterou lze využít pro výpočty softwaru Maxent. Bylo potřeba získat data reprezentující geografický výskyt daného jevu spolu s environmentálními proměnnými. Postupně vznikajícím cílem diplomové práce bylo vytvořit model umožňující hodnotit polohovou přesnost příspěvků přicházejících do aplikace ZmapujTo, které jsou reportovány širokou veřejností (princip crowdsourcingu). Pro tyto účely byly selektovány datové vrstvy:

- ZmapujTo – geografický výskyt daných jevů,
- IPR Praha – současný stav využití území,
– funkční plochy.

4.1.1 Úprava dat ze ZmapujTo

Data z aplikace ZmapujTo jsou publikována ve formátu GeoJSON. Data byla zobrazena v internetovém prohlížeči Mozilla Firefox a manuálně překopírována do nového poznámkového bloku. U takto vzniklého textového dokumentu s koncovkou *.txt* bylo potřeba ručně přepsat koncovku souboru na *.geojson*. Přepis byl proveden v aplikaci Salamander.

Pro převod geodat ze strojově čitelného formátu GeoJSON do formátu Esri shapefile byl zvolen software QGIS 2.6.0. V tomto programu byla zvolena možnost „Přidat vektorovou vrstvu“ a pro zadání zdroje dat byla zvolena možnost „procházet“ – pro zobrazení vyhledávaného souboru bylo nutné zvolit formát souboru GeoJSON (*.geojson *GEOJSON). Jakmile byla takto otevřená vektorová vrstva zobrazena v hlavním grafickém poli, bylo možné vrstvu uložit do formátu SHP (kliknutím pravého tlačítka myši, ve sloupci „*vrstvy*“ na vybranou vrstvu a zvolit možnost „uložit jako“ a zvolit formát SHP).

Další fáze přípravy těchto dat již probíhala v softwaru ArcGIS for Desktop. Na základě konzultace bylo zvoleno, že samotný model bude hodnotit pouze polohovou přesnost reportů, které zaznamenávají černé skládky nad 50 kg. Originální data, která byla v rámci atributového výběru omezena pouze na označení „*report_typ = dump, size = big*“ a která byla prostorově selektována pro území hlavního města Prahy, obsahovala k 1. 12. 2015 1208 záznamů.

Aby byla zajištěna naprostá správnost vstupních dat, bylo přistoupeno k manuální kontrole jednotlivých reportů. Každý z těchto reportů byl podle fotografie kontrolován, zda je zařazen do správné kategorie. Kategorii je nutné v tomto případě chápat jako

správné atributové zařazení jednotlivých příspěvků. To znamená, že nedojde k nesprávnému označení černé skládky namísto drobného nepořádku, autovraku a podobně. V případě, že byla „černá skládka = dump“ zvolena správně, došlo v další fázi ke kontrole atributu velikosti podle fotografie, která je nutnou součástí každého reportu. Za „velkou černou skládku“ je považována každá skládka s více než 50 kilogramy odpadu. Tento atribut je z fotografie obtížné určit, proto byl v nejasných případech report odebrán. Po takto provedené kontrole obsahovala prostorová data 686 záznamů, u kterých bylo možné jasně určit, že se jedná o velké černé skládky.

681	Point	12830	http://www.zmapujto.cz/issue/128306	dump
682	Point	12033	http://www.zmapujto.cz/issue/120335	dump
683	Point	12865	http://www.zmapujto.cz/issue/128655	dump
684	Point	12316	http://www.zmapujto.cz/issue/123167	dump
685	Point	13216	http://www.zmapujto.cz/issue/132165	dump

Navigation: 1 (0 out of 686 Selected)

Obrázek 4.1 Atributová tabulka po manuální kontrole záznamů (686).
autor: J. Vrkoč, 2016



Obrázek 4.2 Ukázka dvou rozdílných velikostí se shodným označením „dump; size = big“.
autor: J. Vrkoč, 2016

V dalším kroku bylo potřeba vypočítat souřadnice X a Y pomocí nástroje „Data Management tools → Features → Add XY Coordinates“. U jednotlivých záznamů došlo k výpočtu souřadnic. Pro převod souboru ve formátu Esri shapefile do formátu CSV byl využit skript, který je součástí SDM toolboxu. Byl využit skript „SDM tools → Basic tools → Table & Shapefile tools → Shapefile to CSV“. U tohoto skriptu je volitelný vstup, umístění výstupu a název výstupu. Další úpravy byly realizovány v softwaru Microsoft Office 365, přesněji v Microsoft Excel (tabulkový editor). Pro další práci v toolboxu SDM bylo nutné zachovat přesnou strukturu tabulky. Výsledná tabulka musela být ve formátu: *species = dump; dd long = zeměpisná délka; dd lat = zeměpisná šířka*. Je nutné striktně dodržet uvedenou strukturu; v jiném případě by použité skripty a nástroje nebyly funkční. Ostatní sloupce tabulky byly odmazány. Pro zpětné převedení tabulky z CSV formátu do formátu Esri shapefile byl použit opět skript z „SDM tools → Basic tools → Table & Shapefile tools → CSV, TXT or XLS to Shapefile“. Ukázka výsledného formátu tabulky je uvedena v obrázku 4.3.

	FID	Shape *	species	dd long	dd lat
▶	0	Point	dump	14,506373	50,006112
	1	Point	dump	14,434265	49,975051
	2	Point	dump	14,488511	50,086317
	3	Point	dump	14,404996	50,0021
	4	Point	dump	14,547256	50,088322
	5	Point	dump	14,334371	50,03383
	6	Point	dump	14,435391	50,148652
	7	Point	dump	14,376133	50,063694

Obrázek 4.3 Nutné striktní dodržení názvů jednotlivých sloupců.
autor: J. Vrkoč, 2016

U takto předpřipravených dat lze vysledovat shluky bodů, které reprezentují černé skládky, a to v oblasti městské části Prahy 10. Děje se tak díky tomu, že Městská policie Prahy 10 spolupracuje se ZmapujTo a během obchůzek městští policisté hlásí jednotlivé podněty (přeplněné koše, poškozené lavičky, černé skládky aj.) pomocí smartphonů. Touto vzájemnou spoluprací může vzniknout mnohem větší množství reportů než v jiných částech hlavního města Prahy. V poslední části přípravy dat ze ZmapujTo bylo nutné odstranit vzájemně blízké body. K tomuto kroku byl použit nástroj, který je součástí toolboxu SDM „SDM tools → Universal tools → Spatially Rarefy Occurrence Data for SDMs → Spatial Rarefy Occurrence Data (a.k.a spatial filter occurrence data)“. Tento nástroj umožňuje odstranit vzájemně blízké body. Na základě expertního odhadu a konzultace s vedoucím práce byla zvolena hodnota „Resolution to Rarefy Data = 50 m.“ (obrázek 4.4).

Obrázek 4.4 Odstranění vzájemně blízkých bodů, hodnota 50 metrů.
zdroj: sdmtoolbox.org/technical-info#Uni3

Tímto odstraněním vzájemně blízkých bodů je omezena možnost výskytu duplicitních hlášení (jsou kontrolována i administrátorem aplikace). Znamená to tedy, že jsou odstraněny body reprezentující jednu skládku zaznamenanou z různých pozic. Výsledkem výše uvedených úprav dat je zaznamenání 312 černých skládek, které slouží jako vstup do modelovacího softwaru Maxent.

4.1.2 Příprava vstupních rastrů

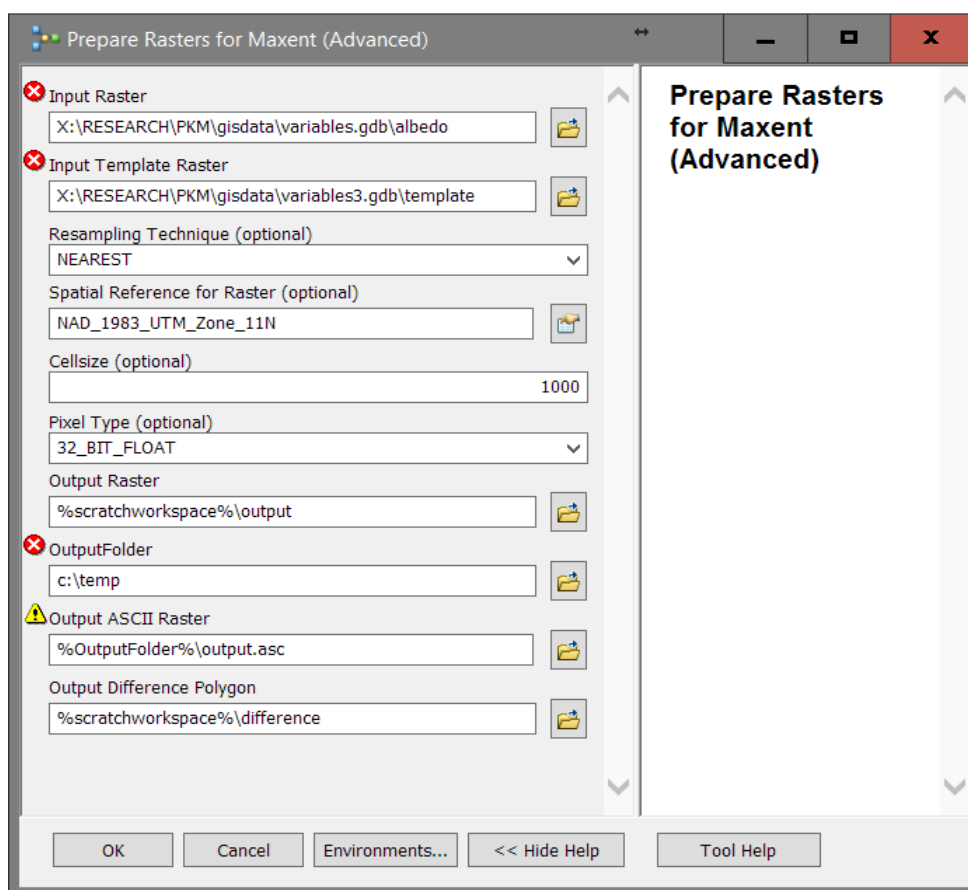
Pro přípravu rastrových vstupních rastrových dat byla zvolena možnost využití dat poskytovaných IPR Praha, který poskytuje data jako „OpenData“. Pro potřeby diplomové práce byly využity vektorové vrstvy současného stavu využití území a funkčních ploch.

Prvotní předpoklad byl založen na využití dvou kategorických rastrů (výše uvedené) a deseti kontinuálních rastrů. U vektorové vrstvy současného stavu využití území byl vždy vybrán určitý typ atributu, např. ulice=VM, les=LRO a další. Vždy došlo k exportu do nové vektorové vrstvy. Následně byl implementován nástroj „*Spatial Analyst tools* → *Distance* → *Euclidean Distance*“. V následných výpočtech se však tento krok ukázal jako nepřesný, respektive chybný, a tudíž nebylo dále s těmito kontinuálními rastry počítáno.

Obě vektorové vrstvy byly z portálu IPR Praha získány v souřadnicovém systému WGS_84. Aby bylo možné zadat při tvorbě rastrových vrstev velikost pixelu v metrech, bylo nutné transformovat vektorové vrstvy do souřadnicového systému WGS_1984_UTM_Zone_33N pomocí toolboxu „*Data Management tools* → *Projections and Transformations* → *Feature* → *Project*“. Obě vektorové vrstvy bylo potřeba převést na rastr pomocí nástroje „*Conversion tools* → *To Raster* → *Feature to Raster*“ s hodnotou výstupní velikosti pixelu 1 metr. V následujícím postupu bylo vyžadováno, aby rastry byly ve formátu *.tiff. Toho lze docílit pomocí skriptu „*Raster to Other Format*“, popřípadě manuálním zadáním koncovky výstupního rastru u nástroje „*Feature to Raster*“.

Další fáze úpravy byla složená ze dvou kroků. V prvním bylo potřeba oba rastry vynásobit hodnotou 100 pomocí nástroje „*Spatial Analyst tools* → *Map Algebra* → *Raster Calculator*“. Jakmile došlo k vynásobení, byl datový typ s vynásobenou hodnotou (dříve Float) převeden na datový typ „*Integer*“ pomocí toolboxu „*Spatial Analyst tools* → *Math* → *Int*“.

V případě, že by v diplomové práci využity kontinuální rastry s různým extentem, bylo by v tomto kroku efektivnější využít toolbox „*PrepareRastersForMaxent*“, který je volně stažitelný na webových stránkách www.arcgis.com.



Obrázek 4.5 Grafické rozhraní toolboxu PrepareRastersForMaxent.

zdroj: <http://www.arcgis.com/home/item.html?id=11bf7e689c92413f8d31933b3e1f56b1>

Jelikož však další postup práce kontinuální rastry nevyužíval, mohlo se přistoupit k tvorbě vlastního modelu (vytvořen pomocí geoprocessingových nástrojů) v „*ModelBuilderu*“ pro přípravu dat, které následně vstupují modelovacího softwaru Maxent.

Předposlední krokem byla transformace rastrů zpět do GCS_WGS_84 pomocí nástroje „*Data Management tools* → *Projections and Transformations* → *Feature* → *Project Raster*“. Poslední částí přípravy rastrových dat byla jejich konverze z rastrového formátu do formátu ASCII pomocí toolboxu „*Conversion tools* → *From Raster* → *Raster to ASCII*“. Výstupem přípravy dat byly dva rastry reprezentující současný stav využití území a funkční plochy. Rastry byly vytvářeny v prostorovém rozlišení 1x1 metr, což zapříčinilo jejich značnou velikost - 3,1 GB a 2,8 GB.

Výsledný model vytvořený pro úpravu dat jako vstupy do „*enviromental layers*“ je značně výpočetně náročný. Výpočty probíhaly zpravidla několik hodin. Model je součástí diplomové práce jako vázaná příloha 1. Současně je model uveden na DVD, které je součástí práce.

4.2 Výpočet v softwaru Maxent

Pro výpočty pravděpodobnosti distribuce skládek v místech, kde nebyly hlášeny, byl použit algoritmus „*maximum entropy = Maxent*“, který je používán pro modelování prostorové distribuce organismů. Označení Maxent pochází z anglického „*maximum entropy*“ označující maximální entropii neboli míru uspořádání. Tento model popisuje ve

svém článku v roce 2016 S. Phillips⁶⁴. Je to statistická metoda, která využívá prezenční data. To je důvodem, proč musela být data pro použití v tomto modelovacím softwaru upravena.

Pro výpočet pravděpodobnosti pomocí softwaru lze využít dva možné přístupy. Prvním přístupem je využití toolboxu SDM, který obsahuje implementovaný algoritmus „*maximum entropy*.“ Jedná se o skript, který se nachází v „*SDM tools* → *MaxEnt tools* → *Modeling with MaxEnt* → *Run MaxEnt: Spatial Jackknifing*“. Tento způsob umožňuje provést složitější analýzy rozšíření druhů.

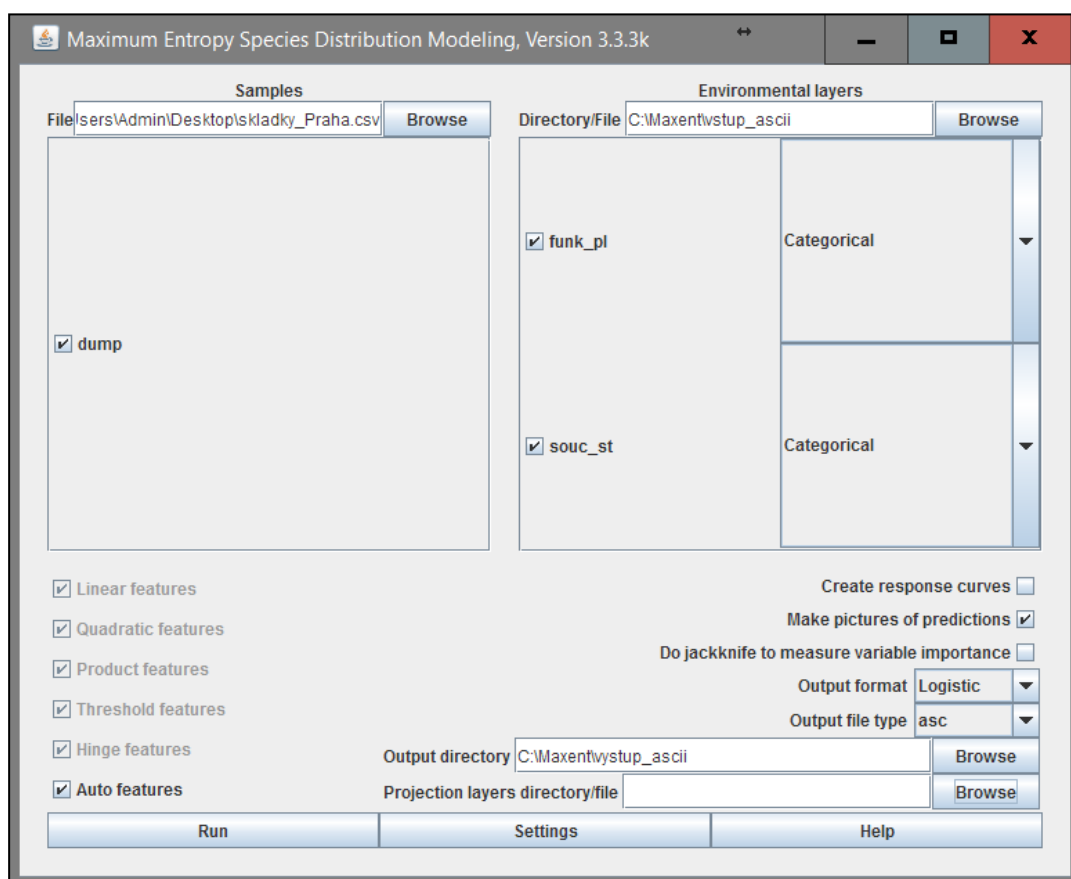
Pro výpočet základního rastru postačoval nástroj „*Maximum Entropy Species Distribution Modeling, Version 3.3.3k*“, který je dostupný pouze přes Javu. Nástroj je volně ke stažení ve formátu *.jar. Grafické rozhraní je v tomto případě výrazně „*user-friendly*“. Výhodou je, že uživatel nemusí mít znalosti práce s GIS softwary od Esri, které umožňují práci s toolboxem.

Pro samotné zahájení výpočtu v modelovacím softwaru je důležité nastavit dva typy vstupních dat. Jednotlivá vstupní data jsou popsána níže:

- Samples – neboli „*vzorek vstupních dat*“. V tomto případě se jednalo o CSV soubor reprezentující geografické umístění černých skládek na území hlavního města Prahy (poloha vyjádřena souřadnicemi X,Y). Jako vstup sloužilo 312 skládek.
- Environmental layers – V této části nastavení byly použity předem upravené ASCII soubory současného stavu využití území a funkčních ploch. Existuje zde možnost zvolit, zda se jedná o rastr kontinuální, či kategorický. V případě těchto rastrů byla zvolena možnost „*kategorický*“.
- Output format – Umožňuje volbu formátu výstupního souboru – byla zvolena možnost *.asc.

V softwaru existuje velké množství nastavení. Pro diplomovou práci však bylo zachováno defaultní nastavení. Jeho součástí je také velmi podrobná nápověda pro práci se softwarem, stejně jako nápověda popisující možnosti nastavení jednotlivých parametrů.

⁶⁴ Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31(2) 161-175.



Obrázek 4.6 Grafické rozhraní softwaru Maxent – nastavení pro výpočet.
autor: J. Vrkoč (2016)

Výstupem výpočtu v Maxentu je soubor ve formátu ASCII (velikost souboru více než 6,5 GB). Pro grafické znázornění výsledků však bylo nutné převést ASCII soubor zpět do rastrové podoby pomocí nástroje „*Conversion tools* → *To Raster* → *ASCII to Raster*“. Výsledný rastr reprezentuje pravděpodobnost výskytů ilegálních (černých) skládek na území hlavního města Prahy.

4.3 Testování a optimalizace

Pro ověření správnosti výsledného modelu bylo nutné provést testování. Ověřování probíhalo na třech odlišných datových sadách:

- 60 náhodných bodů: Tyto body byly vytvořeny automaticky pomocí funkce „*Data Management tools* → *Features Class* → *Create Random Points*“. Při vytváření těchto bodů byla implementována podmínka, že se body budou nacházet ve vzájemné vzdálenosti minimálně 50 metrů.
- 90 expertních bodů: Tyto body byly vytvořeny s předpokladem, že budou umístěny v prostoru do míst, kde je velmi nízká pravděpodobnost, že se zde velká černá skládka vyskytne (popřípadě bude pravděpodobnost nulová, to znamená, že výskyt zde není možný). Příkladem může být střed vodního toku, Staroměstské náměstí, Karlův most, nemocniční zařízení a podobně.
- 52 černých skládek: V tomto případě se jedná o nově reportované černé skládky zaznamenané pomocí aplikace ZmapujTo v rozmezí od 1. 1. 2016 do 20. 3. 2016. Opět byly vybrány pouze skládky, kde množství odpadu přesahuje 50 kilogramů. Tato data byla podrobena, stejně jako vstupní data, vizuální analýze. Opět bylo

kontrolováno přiřazení odpovídající kategorie danému reportu a atributové zařazení velikosti – zda se opravdu jedná o velkou černou skládku. Na základě kontroly byly odstraněny tři reporty z důvodu nesprávného zařazení do kategorie černých skládek, jelikož se ve dvou případech jednalo o přeplněné kontejnery na odpad a ve zbylém případě o odstavený autovrak.

Výsledný rastr byl kontrolován dle ortofota. Jedna z možností, jak získat ortofotomapu, je využití prohlížečské služby WMS z Geoportálu ČÚZK. Pro území hlavního města Prahy lze využít jako zdroj ortofoto snímků IPR Praha, který poskytuje otevřená data. V rámci „*opendat*“ lze získat ortofotomapu Prahy v prostorovém rozlišení 10 centimetrů na pixel. Snímky byly pořízeny v roce v srpnu 2015. Komprimovaný soubor ve formátu *.zip (velikost souboru 12 GB) obsahuje jednotlivé snímky dle kladu listů 1:1000.

Pro kontrolu bylo nutné převést hodnoty pravděpodobnosti výskytu z výsledného rastru (slouží jako vstup) do atributové tabulky jednotlivých bodů (vstupní bodová vrstva). Tato část byla realizována pomocí nástroje „*Spatial Analyst tools* → *Extraction* → *Extract Values to Points*“. Při převodu je v atributové tabulce nově vzniklé bodové vrstvy vytvořen nový sloupec s názvem „*RASTERVALU*“. Jedná se o hodnotu přenesenou ze vstupního rastru.

V poslední řadě bylo nutné stanovit prahovou hodnotu (*threshold* = *práh*). Pro stanovení této hodnoty bylo použito 60 náhodně vytvořených bodů reprezentující černé skládky. Jednotlivé body byly hodnoceny expertním odhadem, který byl realizován autorem. Samotný odhad představoval číselnou hodnotu pravděpodobnosti, s jakou se v dané lokalitě mohla vyskytovat černá skládka. Hodnota každého bodu byla zaznamenána do příslušné buňky atributové tabulky. Jakmile bylo všech šedesát bodů zkontrolováno vizuálně pomocí ortofotomapy, byly do bodového shapefile náhodných bodů přeneseny hodnoty z pravděpodobnostního rastru.

Hodnoty pravděpodobnosti jednotlivých bodů se nacházely na intervalu od 8,2 % po 87,8 %. U bodů s pravděpodobností 30 % docházelo k výraznému lomu. Z 60 bodů bylo 31 bodů nad 30 % a 29 bodů pod 30 %.

V atributové tabulce byl vytvořen nový sloupec. Bodům nad 30% byla přiřazena hodnota 1 a bodům pod 30 % hodnota 0. Následně byly všechny body opět kontrolovány dle ortofotomapy. Z těchto 60 bodů bylo 56 zařazeno správně (z 31 bodů 30 správně; z 29 bodů 26 správně). Přesnost stanoveného prahu byla 93 %, takže lze tuto hodnotu označit za správnou. Po ověření správnosti bylo možné definovat hodnotu 30 % jako prahovou hodnotu (*threshold* = *práh*). Třicet procent je expertně stanovená hodnota prahu, vyjadřující hranici pravděpodobnosti, nad kterou existuje předpoklad, že se zde skládka vyskytuje. Hodnota prahu byla diskutována s Miroslavem Kubáskem a pro potřeby práce byla stanovena jako dostačující.

4.3.1 Testování I

První část testování probíhala na expertně vytvořených bodech, které jsou záměrně umístěny v místech, kde není v reálném světě výskyt těchto bodů reprezentujících černé skládky možný, popřípadě je pravděpodobnost výskytu naprosto minimální. Tyto body byly podrobeny výše uvedenému procesu zjištění pravděpodobnosti. Byl stanoven předpoklad, že všechny body by měly získat hodnotu pravděpodobnosti nižší než 30 % dle stanoveného prahu. Výsledkem ovšem bylo, že z 90 bodů bylo pouze 62 pod 30 % a 28 nad třicetiprocentní hranicí. V tomto případě byla hodnota přesnosti daného rastru 68,8 %.

11	Point	0	23,8945
15	Point	0	25,11164
17	Point	0	25,31646
66	Point	0	28,17649
80	Point	0	30,16858
83	Point	0	32,17649
72	Point	0	37,1436

« 1 » (62 out of 90 Selected)

Obrázek 4.7 Ukázka atributové tabulky u expertních bodů (30% hranice).
autor: J. Vrkoč (2016)

Druhým způsobem ověření správnosti vytvořeného pravděpodobnostního rastru bylo otestovat přesnost na reálných datech, které představovalo 52 nových reportů černých skládek na území hl. města Prahy. V případě těchto 52 bodů reálně nahlášených černých skládek bylo 38 skládek zařazeno správně nad 30 %, zbylých 14 skládek nedosáhlo prahové hodnoty 30 %. Správnost rastru byla stanovena na 73 %.

Prvotním cílem bylo vytvořit nástroj odstraňující polohově nepřesné reporty. V průběhu práce však došlo k přehodnocení cíle. Na základě spolupráce s RNDr. Miroslavem Kubáskem, jakožto autorem aplikace ZmapujTo, bylo zjištěno, že z hlediska polohové nepřesnosti je odstraněno přibližně jedno hlášení z dvou set (příkladem může být nahlášení velké černé skládky uprostřed náměstí). Z tohoto důvodu by prvotní návrh postrádal valné využití. Návrh nástroje byl přehodnocen do podoby poloautomatického nástroje umožňující odesílat reporty s pravděpodobností výskytu vyšší než prahová hodnota automaticky příslušným úřadům. V opačném případě by byla nutná manuální administrace. Cílem se tudíž stalo proces prostorové kontroly co nejvíce poloautomatizovat.

4.3.2 Optimalizace I

Na základě výše uvedených hodnot přesností bylo zřejmé, že existuje prostor pro zpřesnění pravděpodobnostního rastru. Bylo tedy přistoupeno k jeho optimalizaci. Pro optimalizaci byla zvolena vrstva současného stavu využití území. Podle atributu „KOD“ bylo vybráno 66 typů využití území s minimální pravděpodobností výskytu černých skládek na základě expertního odhadu (bytové domy, církevní budovy, nemocnice a další).

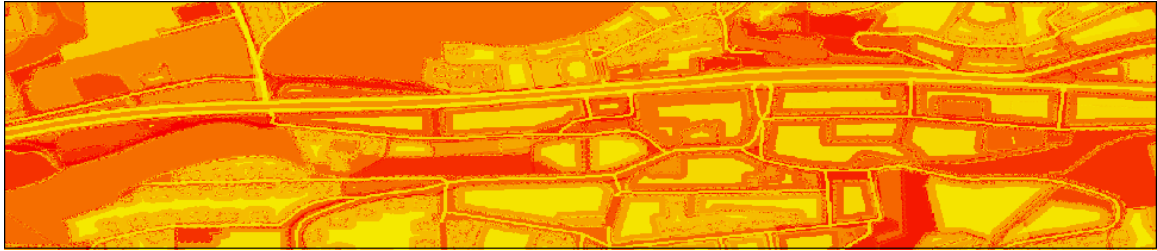
Pro optimalizaci byl vytvořen model v „ModelBuilderu“ (vázaná příloha 2). Model je složen z jednoduchých geoprocessingových nástrojů. V první kroku bylo potřeba vyselektovat z vrstvy současného stavu využití území všechny plochy odpovídající některému z vybraných 66 typů kódu. Následně byl vypočten vnitřní „buffer“ neboli obalová zóna „Analysis tools → Proximity → Buffer“. Vnitřní buffer je vypočítáván stejně jako buffer klasický, avšak s tím rozdílem, že do pole „distance“ je zadávána záporná (minusová) hodnota a do pole „side type = outside_only“. V tomto případě byla zadána hodnota -10 metrů a současně zvolena možnost „dissolve“. Pomocí nástroje „Analysis tools → Overlay → Erase“ byla z vrstvy dříve vytvořené výběrem dle atributu „KOD“ odmazána vrstva vznikla výpočtem vnitřní obalové zóny (vnitřní buffer). V takto nově vzniklé vrstvě byl jako výstup předchozího kroku vytvořen nový sloupec „maxent“, v němž bylo všem buňkám daného sloupce pomocí „Field Calculator“ přiřazena hodnota 5. Tato vrstva byla konvertována pomocí nástroje „Feature to Raster“. Výsledkem je rastr o velikosti pixelu 1x1 metr a hodnotami 5 a NoData. Rastr bylo nutné reklasifikovat pomocí nástroje „3D Analyst tools → Rasters Reclass → Reclassify“.

Hodnota 5 byla zachována a hodnotě NoData byla přiřazena hodnota 1. Finálním krokem je dělení původního výsledného pravděpodobnostního rastru tímto optimalizačním rastrem pomocí nástroje „*Spatial Analyst tools* → *Map Algebra* → *Raster Calculator*“. Ve výsledku výše uvedený postup umožní snížit pravděpodobnost původního rastru pětinašobně v místech, kde má optimalizační rastr hodnotu 5. To znamená, že pokud by původní rastr měl v některém místě hodnotu 100 % (což je nereálné), po provedení optimalizace by měl ve stejném místě hodnotu 20 %. Stejně jako výše uvedené stanovení prahové hodnoty je i tento postup optimalizace založen na expertním odhadu. V praxi se jedná o implementaci zpřesňujících poznatků získaných během kontroly jednotlivých výstupů.

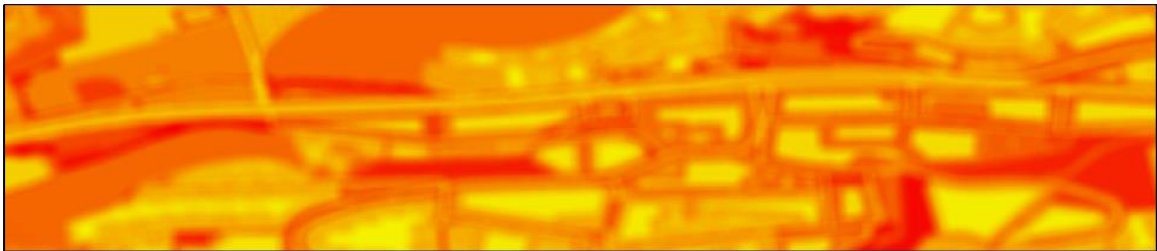
Během zběžné vizuální kontroly byl odhalen problém s možností umístění skládky doprostřed komunikace. Není reálně možné, aby se velká skládka nacházela ve středu komunikace, avšak již optimalizovaný rastr by takovéto umístění skládky ohodnotil pravděpodobností vyšší než 30 %. Proto došlo k druhé optimalizaci. V tomto případě byla optimalizována uliční a komunikační síť. Komunikace byly získány z RÚIAN a převedeny do geodatabáze pomocí nástroje „*VFR Import tool*“. Z vektorové vrstvy současného stavu využití území byly vyexportovány významné komunikace „*KOD = VN (významné ulice a silnice)*“ a běžné ulice „*KOD = VM (ulice a silnice)*“. Vznikly tak dvě nové vektorové vrstvy. Pomocí funkce „*Select by Locations*“ byly z databáze RÚIAN zjištěny významné silnice, které mají průnik (intersect) se silnicemi označenými atributem „*VN*“. Došlo k exportu do nové vektorové vrstvy. Předchozí postup byl realizován pro obě vrstvy reprezentující uliční, respektive komunikační síť. U těchto významných komunikací byla vypočtena obalová zóna (buffer) s poloměrem 3 metry. V případě běžných ulic byla obalová zóna stanovena na 2 metry. V poslední části byly tyto dva rozdílné vektorové shapefile spojeny pomocí funkce „*Analysis tools* → *Overlay* → *Union*“. Na takto vzniklou vektorovou vrstvu byl aplikován optimalizační postup od kroku tvorby nového atributu s hodnotou 5.

4.3.3 Optimalizace II

Finální část optimalizace zahrnovalo řešení možné polohové chyby při hlášení černé skládky při použití lokace pomocí GPS v mobilním zařízení. Možná chyba byla stanovena na hodnotu 10 metrů. Ve výše uvedeném optimalizovaném rastru existovala možnost, že se vedle sebe nacházejí pixely s výrazně odlišnou hodnotou. Bylo proto nutné navrhnout řešení vedoucí k odstranění této chyby. Optimalizovaný rastr byl „zhlazen“ pomocí filtru „*Spatial Analyst tools* → *Neighborhood* → *Focal Statistic*“. Pro výpočet bylo nastaveno sousedství „*neighborhood=circle*“, rádius „*cell=10*“ a typ statistiky „*mean*“ (obrázek 4.9).



Obrázek 4.8 Optimalizovaný rastr bez použití „*Focal Statistics*“.
autor: J. Vrkoč (2016)



Obrázek 4.9 Optimalizovaný rastr zhlazený filtrem „*Focal Statistics*“.
autor: J. Vrkoč (2016)

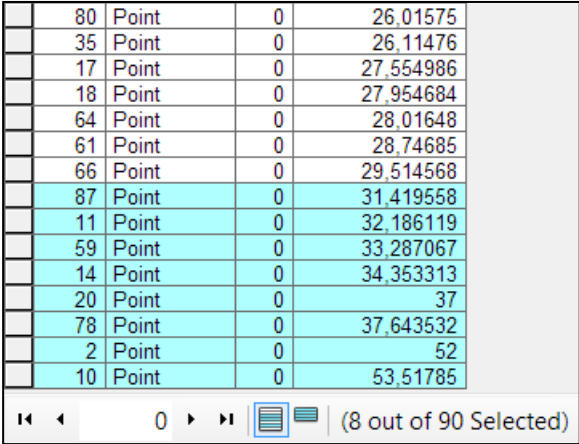
5 VÝSLEDKY

Výsledkem praktické části diplomové práce skládající se z procesu přípravy dat, výpočtů a optimalizace je rastr reprezentující pravděpodobnost výskytu černých skládek nad 50 kilogramů na území hlavního města Prahy. V budoucnu bude sloužit pro ověřování prostorové reálnosti hlášení od občanů, jelikož v současné době je veškerá kontrola prováděna manuálně. Výstupy této diplomové práce budou současně poskytnuty IPR Praha. Na základě těchto dat bude možno detekovat místa s vyšším výskytem černých skládek na území hl. m. Prahy.

5.1 Přesnost optimalizovaného rastru

Po provedení druhé optimalizace následovalo závěrečné testování. Kontrola probíhala stejně jako v předchozích případech, tedy manuálně nad ortofotomapou na základě expertního odhadu. Tento expertní odhad byl následně porovnán s hodnotami převzatými z optimalizovaného rastru.

V první části byla kontrolována pravděpodobnost výskytu 90 expertně vytvořených bodů, které se nachází v místech, kde je výskyt černé skládky nereálný, popřípadě jen velmi málo pravděpodobný. Vychází se z předpokladu, že žádný z těchto bodů by neměl získat hodnotu vyšší než stanovená prahová hodnota 30 %. Z těchto bodů mělo pouze osm více než 30 % pravděpodobnosti výskytu. Ostatních 82 bodů mělo hodnoty pravděpodobnosti nižší než 30 %. Výsledkem tedy je, že 91 % bodů bylo zařazeno správně. V tomto případě byl zaznamenán nárůst oproti neoptimalizovanému rastru o 23 %.



80	Point	0	26,01575
35	Point	0	26,11476
17	Point	0	27,554986
18	Point	0	27,954684
64	Point	0	28,01648
61	Point	0	28,74685
66	Point	0	29,514568
87	Point	0	31,419558
11	Point	0	32,186119
59	Point	0	33,287067
14	Point	0	34,353313
20	Point	0	37
78	Point	0	37,643532
2	Point	0	52
10	Point	0	53,51785

Obrázek 5.1 Ukázka atributové tabulky (výběr hodnot nad 30%)
autor: J. Vrkoč (2016)

Druhá část kontroly byla zaměřena na ověření správnosti na reálných bodech získaných z aplikace ZmapujTo. Bylo získáno 52 bodů reprezentujících velké černé skládky na území hlavního města Prahy. Tyto body byly vizuálně kontrolovány dle fotografie. Výsledkem bylo zařazení 48 bodů jako reálných (to znamená nad 30 %). Čtyři body byly zařazené špatně – což vykazuje 92% přesnost zařazení. Došlo k nárůstu přesnosti o 19 %.

Závěrečnou fází ověření distribučního modelu bylo vyhodnocení reálných dat z aplikace ZmapujTo. Byla získána data obsahující 312 černých skládek. Tyto skládky byly administrátorem vyhodnoceny jako relevantní. Současně proběhla kontrola vizuální dle fotografie, podle nichž se ověřovalo, zda se v daném místě skutečně

vyskytují, popřípadě vyskytovaly, a zda souhlasí atributové zařazení do kategorie „odpady, nečistota“, typ „černá skládka“. Současně musel odpovídat předem zvolený parametr velikosti skládky, to znamená skládka nad 50 kg. Jelikož administrátor tato hlášení povolil k dalšímu zpracování, existuje předpoklad, že by se skládky měly nacházet v místě s vyšší pravděpodobností výskytu než expertně stanovená hranice 30 %.

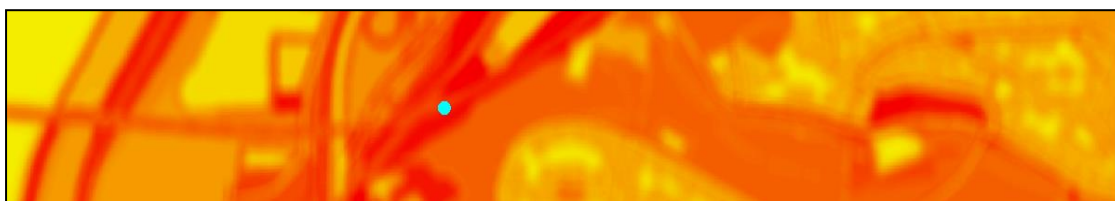
V případě, že by v budoucnu došlo k přijetí navrženého modelu a k implementaci poloautomatického kontrolního nástroje do aplikace ZmapujTo, došlo by tak k výraznému ušetření času administrátora při kontrole. Ušetření času lze doložit na výše uvedeném příkladu 312 nahlášených skládkách, z nichž bylo k automatickému přeposlání hlášení předáno 279 reportů a pouze 31 z nich by bylo potřeba manuálně zkontrolovat. Z toho vyplývá, že automaticky by bylo předáno 90 % hlášení (a pouze každé desáté hlášení by bylo nutno zkontrolovat manuálně).



Fotografie velké černé skládky (nad 50 kg) zaznamenána v aplikaci ZmapujTo.



Umístění reportu (černé skládky) do interaktivní mapy aplikace ZmapujTo.



Zjištění hodnoty pravděpodobnosti výskytu černé skládky v daném bodě reprezentující černou skládku.



FID	Shape *	species	dd long	dd lat	RASTERVALU
180	Point	dump	14.406087	50.026899	54.82019

Výpis hodnot z atributové tabulky (souřadnice daného bodu, hodnota pravděpodobnosti).

Obrázek 5.2 Návrh principu zjištění polohové přesnosti hlášení reportů černých skládek
autor: J. Vrkoč (2016)

Výše uvedený obrázek 4.13 ukazuje jednu z možností principu fungování nástroje pro hodnocení polohové kvality příspěvků do aplikace ZmapujTo. Je důležité si uvědomit, že pravděpodobnostní rastr je vytvořen pouze pro vyhodnocování polohové přesnosti (kvality) příspěvků týkajících se velkých černých skládek. Pro velké černé skládky byla použita specifická sada pravidel (podmínek jejich výskytů), které není možné použít pro jiný typ hlášení. Toto je jedna z nevýhod tohoto typu řešení, protože pro jednotlivé kategorie, respektive typy hlášení by bylo potřeba vytvořit pravděpodobnostní rastry, nehledě na možnost zadávání různých velikostí, jako je tomu v případě hlášení černých skládek.

Součástí diplomové práce jsou dvě mapy. První mapa vyjadřuje průměrnou pravděpodobnost výskytu velkých černých skládek v jednotlivých městských částech hlavního města Prahy (vázaná příloha 3). Hodnoty odpovídající jednotlivým městským částem jsou uvedeny v grafu, který je součástí diplomové práce jako vázaná příloha 4. Druhá mapa reprezentuje pravděpodobnost výskytu černých skládek nad 50 kg na území hlavního města Prahy (vázaná příloha 5).

Z grafu je patrné že městské části *Praha-Přední Kopaniny* a *Praha-Libuš* jsou jediné dvě městské části, které mají průměrnou pravděpodobnost výskytu velkých černých skládek nižší než 30 %. U městské části Přední Kopaniny je tato nízká hodnota (27,16 %) způsobena tím, že je tato městská část z větší části tvořena poli v blízkosti mezinárodního letiště Václava Havla, (bývalého letiště Praha-Ruzyně). Na rozdíl od městské části Přední Kopaniny je městská část Praha-Libuš tvořena z valné většiny zástavbou s rozvinutou silniční sítí. To je také důvodem nízké pravděpodobnosti výskytu velkých černých skládek.

Naopak městskými částmi s nejvyšší pravděpodobností výskytu velkých černých skládek jsou *Praha-Satalice*, *Praha-Běchovice* a *Praha 9 a 16* (více než 40,5 % výskytu). Z hlediska rozlohy se jedná o malé městské části. Existuje předpoklad, že se černé skládky obecně vyskytují v blízkosti železniční tratí a koridorů a v blízkosti přílehlých průmyslových budov a ubytovacích zařízení. V těchto lokalitách dochází ke koncentraci bezdomovců, respektive sociálně slabých a vyloučených rodin. V blízkosti železničních tratí bývají často situovány ubytovny, u nichž se často řeší problematika s čistotou životního prostředí. Dokladem toho může být lokalizace městské části Praha Satalice, která leží na trati Praha-Turnov; podobně Praha-Běchovice leží na trati Praha-Kolín a Praha 16 ležící na trati Praha-Plzeň-Cheb. V případě tří výše uvedených městských částí se jedná o relativně okrajové části Prahy, kde je pravděpodobnost vzniku černé skládky vyšší než v centrální části hlavního města Prahy. V souladu s tím se zde nachází větší zalesněné plochy a méně frekventované silnice, které umožňují vznik velkých černých skládek.

Výsledky práce budou, spolu s výsledným rastrem a modely pro úpravu a optimalizaci dat, poskytnuty Institutu plánování a rozvoje hlavního města Prahy. Pravděpodobnostní rastr lze využít pro predikci míst, kde se černé skládky vyskytují častěji, a díky tomu stanovit místa, kde by mohly být instalovány fotopasti. Současně bude diplomová práce spolu s posterem v digitální podobě poskytnuta IPR Praha.

V rámci diplomové práce byl ve spolupráci Janem Brusem a Miroslavem Kubáskem vypracován článek jako příspěvek na „8th International Congress on Environmental Modelling & Software – iEMSs 2016“. Článek byl odeslán k recenzi.

5.2 Návrh konceptu pro komplexní hodnocení kvality hlášení v rámci aplikace ZmapujTo

Praktickou realizací konceptu by vznikl poloautomatický nástroj přímo propojený s aplikací ZmapujTo umožňující zpracování reportů. V současnosti probíhá zpracování reportů manuálně. Tvorbou poloautomatického nástroje by došlo k výrazné úspoře času administrátora. Úkolem konceptu je umožnit provádění hodnocení kvality z pohledu polohové, atributové a časové přesnosti. Jednotlivé možnosti jsou popsány níže:

Polohová přesnost

Zajistit polohovou přesnost je možné při využití již připraveného pravděpodobnostního rastru, který je součástí diplomové práce. Pokud by tento rastr již nebyl aktuální, je možné využít detailní postup pro jeho tvorbu.

Další možností je využití *dotazování* při zadávání polohy v podobě „*Nacházíte se na místě skládky?*“. V případě volby „*ano*“ by následoval druhý dotaz „*Využíváte GPS mobilního zařízení?*“. Pokud by uživatel zvolil dvakrát možnost ano, mohl by být tento příspěvek brán jako prostorově přesný. V případě využití zadání polohy do interaktivní mapy aplikace ZmapujTo by bylo nutno report zkontrolovat – existuje zde možnost vzniku prostorové chyby.

Atributová přesnost

Zajištění atributové přesnosti je zásadním krokem pro komplexní hodnocení kvality prostorových dat. V případě praktického řešení pro černé skládky nad 50 kg je nutno nejprve zajistit atributovou přesnost v případě určení kategorie, tj., že se jedná o černou skládku a nikoli o přeplněný kontejner, popřípadě autovrak apod. Pro zajištění správné kategorie lze využít pokročilejších informatických řešení v podobě použití algoritmů využívajících neuronové sítě, které dokáží analyzovat fotografie (musí být součástí každého reportu ZmapujTo) a detekovat na něm objekty. Neuronová síť se učí stejně jako lidský mozek. Fotografie k jednotlivým reportům jsou ukládány na „*www.cloudinary.com*“. Každý snímek by musel být popsán např. deseti slovy, co se na něm nachází. Následně by došlo k samotné fázi „učení“. Ani tento systém však není dokonalý a může vykazovat chyby. Hlavní nevýhodou černých skládek je, že se v nich mohou nacházet rozličné předměty - od plastů, kovů, stavební suti, přes bioodpad či pneumatik po spousty dalších objektů. I přes zmíněné nedostatky se ovšem stále jedná o jednu z možností, kterou by bylo možno pro hodnocení polohové přesnosti využít.

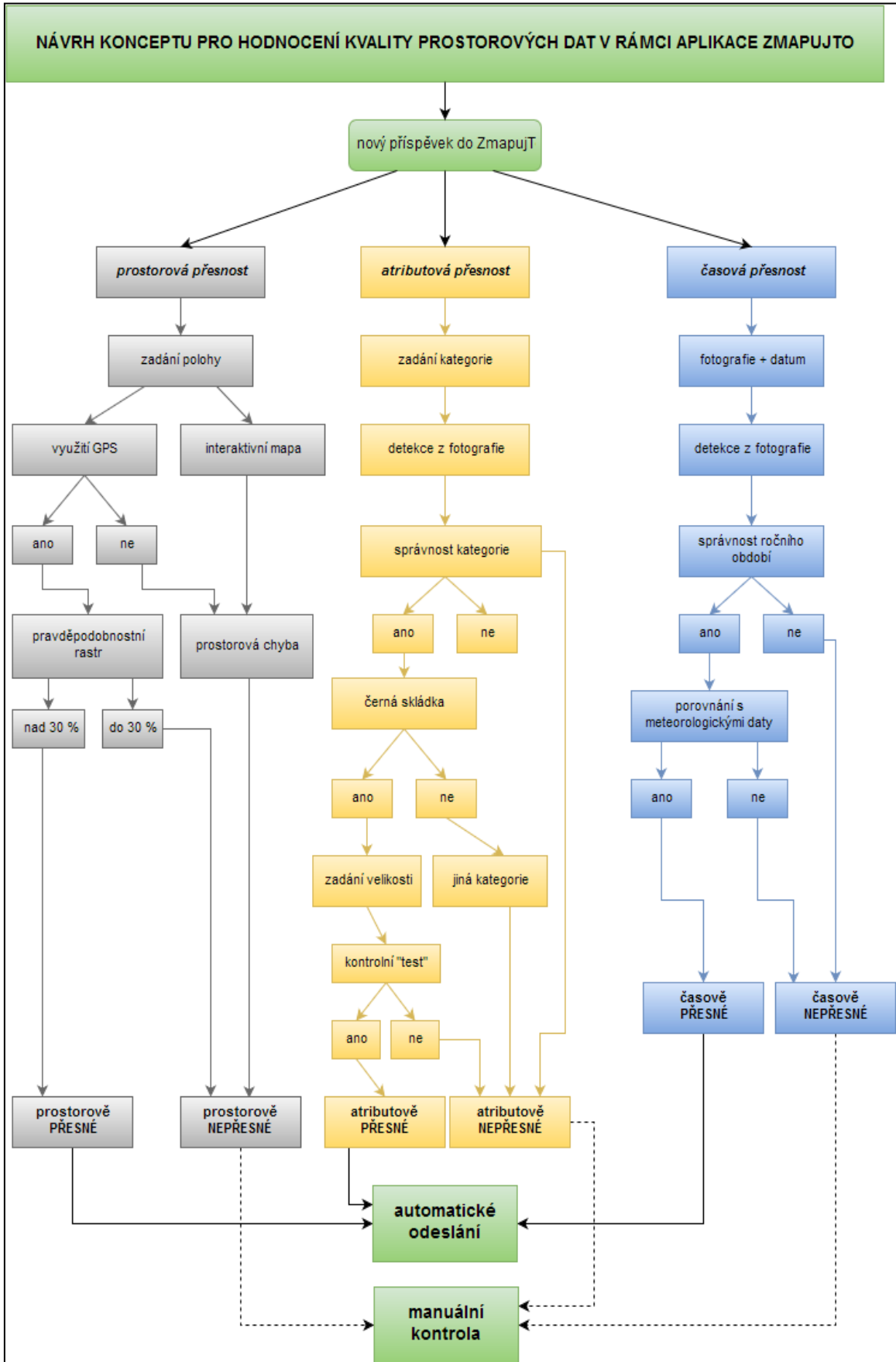
Atributovou přesnost je nutné zajistit také při zadávání velikosti černých skládek. V současnosti je možné určit tři váhové kategorie: malá (do 10 kg), střední (do 50 kg) a velká. V některých případech (současně pro některé uživatele) je zadávání hmotnosti značně problematické. Váhový odhad je dosti subjektivní a ne každý uživatel je váhový interval schopen určit přesně. Při zadávání reportů kategorie „*černá skládka*“ by byl spuštěn krátký test pro stanovení hmotnosti černé skládky na fotografii. Jakmile by uživatel odpověděl v prvních třech testech správně, byl by systémem označen jako „*důvěryhodný*“ a v dalších hlášeních už by test spouštěn nebyl.

Časová přesnost

Časová přesnost je v případě aplikace ZmapujTo relativně potlačena. Jednou z možností zajištění časové přesnosti spočívá v opětovném využití detekce z fotografie. Principem by bylo vyhodnocení fotografií z hlediska ročního období. V případě, že by uživatel zaslal report s fotografií černé skládky, která je pokryta sněhem, respektive ledem, v letním období byl by tento report automaticky označen jako „*časově nedůvěryhodný*“ a byl by přesunut k manuální administraci.

Propojení detekce fotografie s meteorologickými daty je dalším způsobem kontroly časové přesnosti. V případě, že uživatel zašle fotografii v 15:00 s modrou oblohou a slunečným počasím a meteorologická data budou vykazovat ve stejnou dobu pro stejné deštivé počasí, opět bude hlášení označeno jako „*časově nedůvěryhodné*“.

Existují i další možnosti, jak se vypořádat s problematikou hodnocení kvality crowdsourcingových prostorových dat. Výše uvedené návrhy jsou pouze teoretickými nástiny bez hlubšího zkoumání realizovatelného řešení. Schematicky je tento koncept uveden v obrázku 4.14.



Obrázek 5.3 Návrh konceptu pro komplexní hodnocení kvality prostorových dat v aplikaci ZmapujTo
 autor: J. Vrkoč (2016)

6 DISKUZE

Crowdsourcing obecně je velmi náchylný na ochotu komunity pravidelně přispívat. Jedním ze základních pilířů crowdsourcingu je, že přispěvatelé nejsou finančně ohodnoceni, popřípadě jsou ohodnoceni až na úspěšném konci projektu. To je důvod, proč klasické firmy či organizaci mají větší šanci dokončit zadaný projekt i za předpokladů, že dojde k navýšení nákladů.

Už ze samotné podstaty crowdsourcingu vyplývá, že zajistit kvalitu prostorových dat je značně obtížné. U velkých projektů je kvalita zajišťována algoritmicky (příkladem může být OSM). Aplikace ZmapujTo, založená na principu crowdsourcingu, využívá mapování komunitou. V současnosti však není nijak řešeno hodnocení kvality, respektive přesnosti jednotlivých příspěvků.

Jedním z cílů diplomové práce bylo vytvořit nástroj pro hodnocení polohové přesnosti dat získaných z aplikace ZmapujTo. Výsledný model může tvořit základ pro tvorbu poloautomatického nástroje přímo propojeného s aplikací ZmapujTo, který by umožnil zpracování reportů. Pro polohovou kontrolu dat byla stanovena hraniční hodnota (práh), tzv. threshold. V případě, že by pravděpodobnost výskytu byla nižší než stanovená hodnota, došlo by k manuální administraci. V opačném případě by byl report automaticky přeposlán danému oddělení města či obce, které má problematiku černých skládek na starosti. Je důležité si uvědomit, že tento navržený nástroj by řešil pouze polohovou přesnost. To znamená, zda je reálné, že se v daném místě černá skládka vyskytuje, či nevyskytuje. Mezi nevýhody automatického přeposílání patří fakt, že by při něm nedocházelo ke kontrole atributové přesnosti, zda lidé správně určí černou skládku a nikoliv drobný nepořádek, popřípadě přeplněný kontejner. Současně koncept naráží na problém správné lokalizace skládky. Proto je nutno propojit kontrolu polohové a atributové přesnosti do jednoho unitárního kontrolního nástroje.

V procesu zpracování dat lze vysledovat fakt, že bylo získáno 1208 skládek z aplikace ZmapujTo, ale do výpočtu vstoupilo pouze 312 skládek. Nizký počet je způsoben nejen nesprávným zařazením do kategorie černá skládka (často zaměněno s přeplněným kontejnerem), tak atributovou nepřesností v podobě zadávání velikosti černé skládky (v případě hmotnosti). Další výrazným zásahem do originálních dat byl výpočet autokorelace neboli odmazání vzájemně blízkých bodů. Nastavena byla na hodnotu 50 metrů. V tomto případě existuje možnost nastavení odlišné hodnoty podpořené statistickým výpočtem optimální vzdálenosti. Zvýšení současné 90% přesnosti výsledného pravděpodobnostního rastru by mohlo být docíleno vyšším počtem vstupních bodů.

Jednou ze vstupních environmentálních vrstev byla vrstva „*funkční plochy*“, která je součástí územního plánu. Bylo by však nutné při každé změně této vrstvy provést nový výpočet, což by bylo značně časově náročné.

Při výpočtu v softwaru Maxent je naprosto nezbytné, aby všechny vstupní environmentální proměnné (v tomto případě rastry) měly naprosto shodný extent. V případě odlišných extentů nebude výpočet umožněn. Z pravidla bývají rozdílné hodnoty extentu až na posledních třech místech za desetinnou čárkou.

Prvotní předpoklad pro výpočet pravděpodobnostního rastru počítal s využitím kontinuálních rastrů vypočtené pomocí nástroje „*euclidian distances*“ (vzdálenost od vodních toků, garáží, uliční sítě a podobně). V tomto případě však výsledný rastr vykazoval vysoké nepřesnosti, a proto bylo od tohoto postupu upuštěno.

Určování polohy pomocí smartphonu bývá v mnoha případech značně problematické, především, jedná-li se o hustou zástavbu a zalesněný terén. Poloha

může být lokalizována s chybou ± 10 metrů. Příkladem této polohové chyby může být umístění některých černých skládek uprostřed silnice, ve vodním toku, případně uprostřed budovy. Výsledný model se možnou polohovou chybu snaží eliminovat s využitím fokálního kruhového filtru s poloměrem 10 metrů.

Umístění uprostřed budovy či domu nicméně nemusí vždy znamenat chybu v datech. V případě, že mobilní telefon nedisponuje připojením pomocí datových paketů, umožňuje aplikace ZmapujTo report zaznamenat do paměti telefonu a následně odeslat při připojení na WiFi. To však současně zapřičiňuje, že se zaměří poloha, odkud byl daný report do aplikace odeslán. Tato chyba lze odstranit manuálním určením polohy přímo do interaktivní mapy, což daná aplikace umožňuje také. Druhou možností je implementace jednoduché otázky „*Nacházíte se na daném místě?*“.

Výpočet pravděpodobnostního rastru výskytu černých skládek pro velká města ČR, popřípadě pro celou Českou republiku by byl značně problematický. Ne všechna města sdílí stejný postoj přístupu k datům jako hlavní město Praha, které poskytuje vybraná data jako OpenData. Diplomová práce je postavena striktně na využití OpenDat, jak už ze ZmapujTo, tak z IPR Praha. Dalším problémem by byla značná výpočetní náročnost projektu. Konverze a výpočty pro hl. m. Prahu trvaly v některých případech osm a více hodin.

Samotný model pravděpodobnosti výskytu černých skládek nad 50 kg může sloužit jednotlivým městským částem hlavního města Prahy pro detekci míst, kde je potenciálně zvýšená pravděpodobnost, že se v daném místě černá skládka vyskytne. Za tímto účelem budou výsledná data poskytnuta do IPR Praha.

V současnosti ZmapujTo poskytuje data ve strojově zpracovatelném formátu GeoJSON. Aktualizace je realizována pomocí „*cache*“, která se každých 24 hodin aktualizuje. Aplikace ZmapujTo umožňuje filtrování dat podle kategorií. Stáhnutí dat je vždy realizováno v podobě kompletních dat. Vždy je potřeba si požadovaná data vyselektovat. Toto je jedna z možných oblastí, která by mohla být v aplikaci ZmapujTo vylepšena.

Existují však případy, ve kterých kvalita prostorových dat není na prvním místě a může být na úkor rychlosti získávání dat. Crowdsourcingové mobilní aplikace lze využít pro mapování krizových situací (povodně, lesní požáry) a válečných konfliktů (konflikt v Sýrii). V tomto případě je důležitá především aktuálnost a rychlost získávání prostorových dat. Potenciál crowdsourcingu v geoinformatice je značný, avšak je nutné zvážit, pro jaký typ mapování bude tento princip využit.

Pro kvalitu crowdsourcingových prostorových dat v současné době neexistuje vhodný standard. Proto bývají pro hodnocení kvality prostorových dat nejčastěji využívány běžné ISO standardy. Publikace zabývající se crowdsourcingem a jeho využitím, respektive zajištěním kvality takto získávaných dat je v zahraničí velké množství. V České republice je problematika crowdsourcingu řešena pouze okrajově.

Až reálná implementace pravděpodobnostního rastru do nástroje hodnotícího polohovou přesnost by ukázala jeho skutečnou využitelnost při praktickém řešení. Využitím tohoto nástroje by došlo k výrazné úspoře času administrátora při kontrole jednotlivých příspěvků.

7 ZÁVĚR

Crowdsourcing je v posledních letech často prosazovanou metodou v případech mobilního mapování a sběru prostorových dat. Mohutněji se rozvinul až v posledním desetiletí díky narůstajícímu počtu lidí disponujících mobilními zařízeními s přístupem k internetu. V současné době vlastní mnoho lidí chytrý telefon (smartphone) s datovým připojením k internetu. Nejsou tedy závislí pouze na připojení k internetu skrze WiFi. Toto je jedním z nejdůležitějších předpokladů pro rozvoj mobilních aplikací, které jsou obrovským, v současné však době stále nedoceneným, zdrojem prostorových dat založených na principu crowdsourcingu.

U běžných prostorových dat je hodnocení kvality prakticky obstaráno. V České republice jsou pro hodnocení využívány nejčastěji ISO standardy. V současnosti však pro prostorová data vzniklá crowdsourcingem dosud neexistuje specifický standard, který by hodnotil jejich kvalitu. Využití běžných standardů je problematické z toho důvodu, že u takovýchto dat nedochází k fázi standardizace.

V rámci diplomové práce byl navržen postup možného ověřování polohové přesnosti na reálných datech v rámci aplikace ZmapujTo. Jedním z cílů práce bylo striktní využití OpenDat. Výsledná data byla poskytnuta IPR Praha a ZmapujTo. Ověření navrhovaného konceptu bylo prakticky realizováno na případové studii velkých černých skládek (nad 50 kg) pro území hlavního města Prahy.

Systém pro hodnocení polohové přesnosti je pouze jednou z částí komplexního nástroje pro hodnocení kvality reportů přijímaných aplikací ZmapujTo. Princip ověření polohové přesnosti je založen na tvorbě rastru reprezentujícího pravděpodobnost výskytu velkých černých skládek na území hl. m. Prahy. Prostorové rozlišení rastru je 1x1 metr. Správnost, s jakou pravděpodobnostní rastr vyhodnotí polohovou přesnost jednotlivých reportů, je 90 %.

Druhou nedílnou součástí diplomové práce je teoretický návrh konceptu pro komplexní hodnocení kvality prostorových dat v rámci aplikace ZmapujTo. Návrh pro hodnocení polohové přesnosti využívá pravděpodobnostní rastr. Pro hodnocení atributové a časové přesnosti jsou navrženy možnosti, jak lze tuto problematiku řešit. Atributová, neboli tematická přesnost odpovídá přiřazení odpovídající kategorie k reportu (např. černá skládka namísto přeplněného kontejneru), což je v této aplikaci naprosto stěžejní. V kategorii černých skládek umožňuje navíc aplikace zadat také velikost: malá (do 10 kg); střední (do 50 kg); velká. Toto hodnocení pomocí stanovení hmotnosti černé skládky je velmi subjektivní – odhadnout velikost může být pro uživatele značně problematické. Součástí je také možnost způsobu řešení časové přesnosti při využití detekce objektů z fotografie. Výše uvedený návrh testování atributové a časové přesnosti však nebyl ověřován na reálných datech a je součástí diplomové práce v podobě teoretické schématu.

V případě zavedení nástroje do praxe existuje jistý předpoklad, že velkých černých skládek by takto bylo automaticky vyhodnoceno a předáno úřadům až 90 %. Výsledkem by bylo výrazné ušetření času administrátora při kontrole jednotlivých reportů, jelikož doposud byly všechny reporty kontrolovány manuálně.

O výstup řešení projevil zájem také IPR Praha. Mapa pravděpodobnosti výskytu černých skládek může sloužit jako predikce výskytu. Za využití mapy bude IPR Praha moci detekovat místa, kde se skládky vyskytují častěji než jinde na území hlavního města Prahy.

Na rozdíl od mohutných aplikací typu OSM, kde je hodnocení kvality založeno na základě složitých algoritmů, je tento způsob založen striktně na řešení s využitím GIS. Pro tento typ úloh není využívání GIS zcela běžné, proto bylo jedním z hlavních úkolů diplomové práce představit jednu z možností, jak GIS v této problematice úspěšně využít. Tento nástroj je typickým příkladem využití jednodušších analytických nástrojů pro tvorbu komplexního nástroje pro hodnocení polohové přesnosti dat.

Úspěšným propojením crowdsourcingu a GIS se otevírá nová, ne zcela známá, oblast geoinformatiky. Je důležité si uvědomit, pro jaký typ mapování budou aplikace využívající principu crowdsourcingu používány. Není vždy nutně potřeba získávat data vysoké kvality na úkor například rychlosti mapování. V propojení crowdsourcingu a geoinformatiky lze vysledovat značný potenciál pro budoucí využití.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

AGENTURA OCHRANY PŘÍRODY A KRAJINY ČR, Jan Zárybnický. BioLog [online]. Praha, 2015 [cit. 2015-06-07]. Dostupné z: <http://biolog.nature.cz/cz/Article/AboutApp#>

ArcGIS for Desktop: GIS začíná zde. ARCDATA PRAHA [online]. 2016 [cit. 2016-03-30]. Dostupné z: <https://www.arcdata.cz/produkty/arcgis/desktopovy-gis/arcmap>

BRUS, Jan. Vizualizace nejistoty v environmentálních studiích. Olomouc, 2013. disertační práce. 144s. UNIVERZITA PALACKÉHO V OLOMOUCI. Přírodovědecká fakulta

CC BY-NC-ND 3.0 CZ: Uveďte původ-Neužívejte komerčně-Nezpracovávejte 3.0 Česká republika. Creative Commons: Keep the internet creative, free and open [online]. 2015 [cit. 2016-03-30]. Dostupné z: <https://creativecommons.org/licenses/by-nc-nd/3.0/cz/>

CC BY-NC-SA 4.0: Uveďte původ-Neužívejte dílo komerčně-Zachovejte licenci 4.0 Mezinárodní. Creative Commons: Keep the internet creative, free and open [online]. 2015 [cit. 2016-03-30]. Dostupné z: <https://creativecommons.org/licenses/by-sa/4.0/deed.cs>

CHO, George. Some legal concerns with the use of crowd-sourced Geospatial Information [online]. 2014, 16s. [cit. 2015-09-15]. Dostupné z: <http://iopscience.iop.org/article/10.1088/17551315/20/1/0120401/jsessionid=DFE27F6DCDDADA>

Crowdsourcing: síla davu: Waze [online]. Olomouc: GISportal.cz, 2011 [cit. 2016-03-17]. Dostupné z: www.gisportal.cz/2011/03/crowdsourcing-sila-davu/

Crowdsourcing: síla davu: Ushahidi [online]. Olomouc: GISportal.cz, 2011 [cit. 2016-03-17]. Dostupné z: www.gisportal.cz/2011/03/crowdsourcing-sila-davu/

DAR SARMA, Akash, Aditya G. PARAMESWARAN a Jennifer WIDOM. *Globally Optimal Crowdsourcing Quality Management* [online]. 2015, 17s. [cit. 2015-09-15]. Dostupné z: <http://ilpubs.stanford.edu:8090/1111/>

HEER, Jeffrey a Michael BOSTOCK. *Crowdsourcing graphical perception:: using mechanical turk to assess visualization design* [online]. Stanford University, 2010, 10s. [cit. 2015-06-06]. ISBN 978-1-60558-929-9. Dostupné z: [http://dl.acm.org/citation.cfm?id=1753357/](http://dl.acm.org/citation.cfm?id=1753357)

HINTNAUS, Ivo. GISPORTAL.CZ. DejTip, Výmoly, Citysources: crowdsourcingové geoaplikace [online]. 2013 [cit. 2015-06-07]. Dostupné z: <http://www.gisportal.cz/2013/11/dejtip-vymoly-citysourcedcrowdsourcing-geoaplikace/>

HOWE, J. Crowdsourcing:: How the power of the crowd is driving the future of business [online]. Crown business, 2008, 320 s. [cit. 2016-20-03].

Humanitarian Open Street Map Team: Ongoing projects [online]. 2015 [cit. 2016-03-17]. Dostupné z: <https://hotosm.org/projects/ongoing>

IDRIS, N. H., M.J. JACKSON a M.H.I. ISHAK. A conceptual model of the automated credibility assessment of the volunteered geographic information [online]. 2014, 7s., [cit. 2015-09-15]. Dostupné z: <http://iopscience.iop.org/article/10.1088/17551315/18/1/012070/meta>

ISO. ISO 19115:2003 Geographic information - Metadata. 2003b.

ISO. ISO 19131:2007 Geographic information - Data product specifications. 2007.

ISO. ISO/TS 19138:2006 Geographic information - Data quality measures. 2006.

ISO. ISO/DIS 19157 Geographic information - Data quality 2013.

JÖRG STARK, Hans. *Quality assurance of crowdsourced geocoded address-data within OpenAddresses: Concepts and Implementation*. Centre for GeoInformatics, Salzburg University, 2010, 158s., Master Thesis. Centre for GeoInformatics, Salzburg University.

KRESSE, W., FADAIE, K. ISO standards for geographic information. Berlin; New York: Springer, 2004. ISBN 978-3-540-20130-4.

Lucidchart: Our flowchart maker works the way you do [online]. Lucid Software Inc., 2015 [cit. 2016-03-31]. Dostupné z: <https://www.lucidchart.com/>

Lucidpress: AN ONLINE POSTER MAKER FOR EVERYONE [online]. Lucid Software Inc., 2015 [cit. 2016-03-31]. Dostupné z: <https://www.lucidpress.com/>

McCULLAGH, Michael a Mike JACKSON. CROWDSOURCED MAPPING – LETTING AMATEURS INTO THE TEMPLE? [online]. 2013, 34s.. [cit. 2015-09-15]. Dostupné z: <http://www.int-arch-photogramm-remote-sens-spatialinfsci.net/XLW1/399/2013/isprsarch-ive-s-XL-1-W1-399-2013.html>

OpenStreetMap:OSM [online]. 2014 [cit. 2015-06-07]. Dostupné z: <https://www.openstreetmap.org/about>

PÁNEK, Jiří a kol. GeoParticipace: Jak používat prostorové nástroje v rozhodování o lokalitách, ve kterých žijeme?. 1.vyd. Olomouc: Univerzita Palackého v Olomouci, 2014, 75s. 978-80-244-4359-1

PARKER, C.J., MAY, A. and MITCHELL, V., 2012. Using VGI to enhance user judgements of quality and authority. IN: Geographical Information Science Research UK (GISRUK) 2012, 9s., Conference Proceedings.

Phillips, S., Dudik, M. & Schapire, R., 2010, "Maxent Software, ver. 3.3.3e". [online]. 2015 [cit. 2016-03-30]. Dostupné z: <https://www.cs.princeton.edu/~schapire/maxent/>

Plznito [online]. Plzeň: Správa informačních technologií města Plzně, 2016 [cit. 2016-03-17]. Dostupné z: www.plznito.cz; www.plzni.to

Pomozte najít malajsijský boeing [online]. Olomouc: GISportal.cz, 2014 [cit. 2016-03-17]. Dostupné z: <http://www.gisportal.cz/2014/03/pomozte-najit-malajsijsky-boeing/>

PRAŽÁK, J.: Terminologický slovník zeměměřičství a katastru nemovitostí [online]. Zdiaby Výzkumný ústav geodetický, topografický a kartografický, Terminologická komise Českého úřadu zeměměřického a katastrálního, 2005 - 2013, [cit. 12-12-2012]. Dostupné na WWW:<<http://www.vugtk.cz/slovník/index.php>>.

Pražská otevřená data: Co jsou otevřená data. Institut plánování a rozvoje hlavního města Prahy: Geografická data Prahy na jednom místě [online]. Praha, 2015 [cit. 2016-03-30]. Dostupné z: <http://www.geoportalpraha.cz/cs/clanek/271/prazska-otevrena-data#.Vvu1d0dBQuM>

Prepare Rasters for Maxent Tool for ArcGIS 10.1 [online]. 2015 [cit. 2016-03-30]. Dostupné z: <http://www.arcgis.com/home/item.html?id=11bf7e689c92413f8d31933b3e1f56b1>

PSPad: freeware editor [online]. Slavkov u Brna: Jan Fiala, 2015 [cit. 2016-03-31]. Dostupné z: <http://www.pspad.com/cz/>

QGIS: A Free and Open Source Geographic Information System [online]. 2016 [cit. 2016-03-30]. Dostupné z: <http://www.qgis.org/en/docs/index.html>

RICE, Matthew T a kol. *Crowdsourced Geospatial Data: A report on the emerging phenomena of crowdsourced and user-generated geospatial data* [online]. 1. George Mason University, 2012, 147s.[cit. 2015-06-05]. Dostupné z: <http://www.dtic.mil/dtic/tr/fulltext/u2/a576607.pdf>

SHI, W. Principles of modeling uncertainties in spatial data and spatial analyses. Boca Raton: CRC Press/Taylor & Francis, 2010. ISBN 978-14-20059-27-4

SDM toolbox: Home [online]. Duke University, USA, 2015 [cit. 2016-03-30]. Dostupné z: <http://sdmtoolbox.org/>

Syria Tracker: HUMANITARIAN TRACKER [online]. 2011 [cit. 2015-06-07]. Dostupné z: <http://www.humanitariantracker.org/#!/syria-tracker/cj00>

ŠRÁMEK, Petr. *Hodnocení kvality prostorových dat pro zvolené typy prostorových analýz*. Pardubice, 2011. Diplomová práce. Univerzita Pardubice. 112s. Vedoucí práce Doc. Ing. Jitka Komárková, Ph.D.

TRYHUBOVÁ, P. *Principy INSPIRE a standardizace ve vazbě na data ZABAGED R*. 2013. Disertační práce. České vysoké učení technické v Praze. Fakulta stavební, Katedra mapování a kartografie

van EXEL, M., E. DIAS a S. FRUIJTIER., 2012, The impact of crowdsourcing on spatial data quality indicators [online]. [cit. 2015-09-15], 4s., Dostupné z: http://www.researchgate.net/profile/Eduardo_Dias8/publication/267398729_The_impact_of_crowdsourcing_on_spatial_data_quality_indicators/links/546f49af0cf216f8cfa9d247.pdf

Věstník: Úřadu pro technickou normalizaci, metrologii a státní zkušebnictví [online]. 2015(5), 68 [cit. 2016-04-02]. Dostupné z: <http://www.technickenormy.cz/publicdoc/vestnik-05-15.pdf>

Windows Central: What is Microsoft Office 365? [online]. Windows central - Richard Devine, 2015 [cit. 2016-03-30]. Dostupné z: <http://www.windowscentral.com/what-microsoft-office-365>

XIA, Jingfeng. Metrics to Measure Open Geospatial Data Quality. *Issues in Science and Technology Librarianship* [online]. Indiana University, 2012, 9s. [cit. 2016-03-16]. DOI: 10.5062/F4B85627. Dostupné z: <http://www.istl.org/12-winter/article1.html>

PŘÍLOHY

SEZNAM PŘÍLOH

Vázané přílohy:

- Příloha 1 Grafické znázornění modelu pro přípravu dat pro software Maxent
- Příloha 2 Grafické znázornění modelu pro optimalizaci
- Příloha 3 Mapa průměrné pravděpodobnosti výskytu černých skládek v městských částech hlavního města Prahy
- Příloha 4 Graf pravděpodobnosti výskytu černých skládek v městských částech hlavního města Prahy
- Příloha 5 Mapa pravděpodobnosti výskytu černých skládek nad 50 kg v hlavním městě Praha

Volné přílohy

- Příloha 6 Poster
- Příloha 7 DVD

Popis struktury DVD

Adresáře:

Metadata

Přílohy

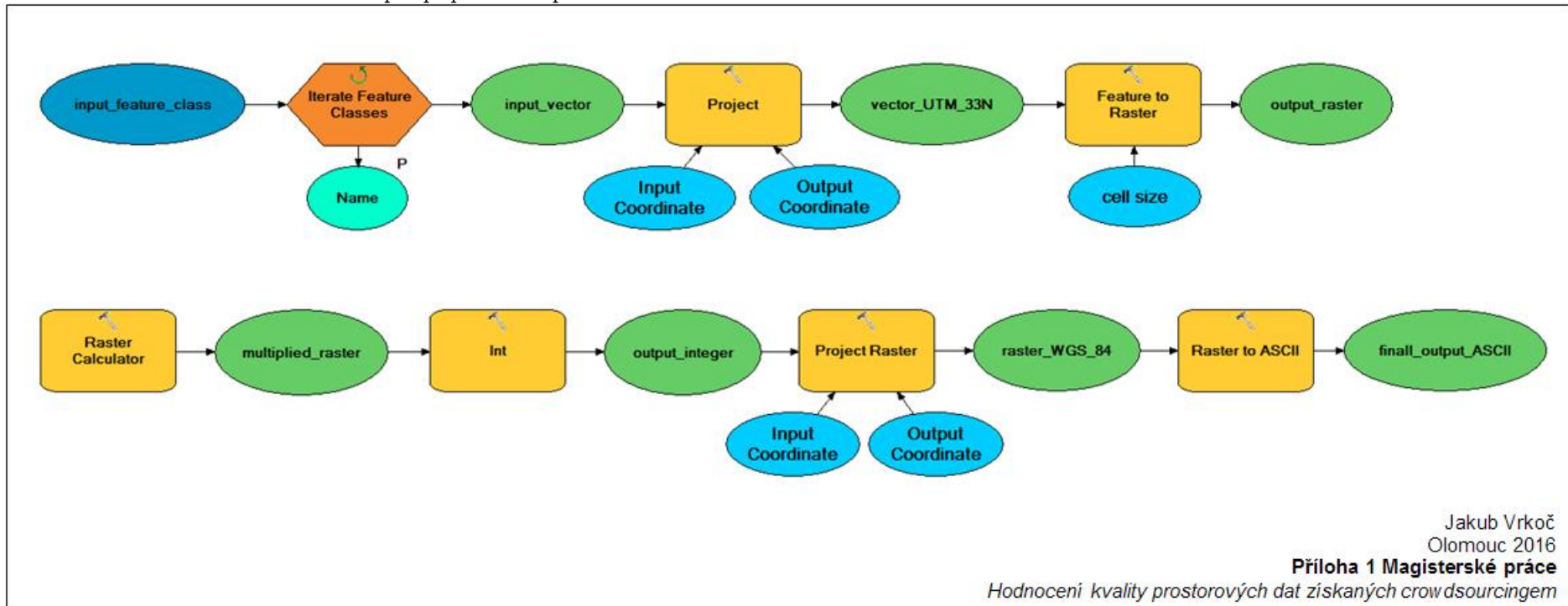
Text_prace

Vstupni_Data

Vystupni_Data

WEB

Příloha 1 Grafické znázornění modelu pro přípravu dat pro software Maxent

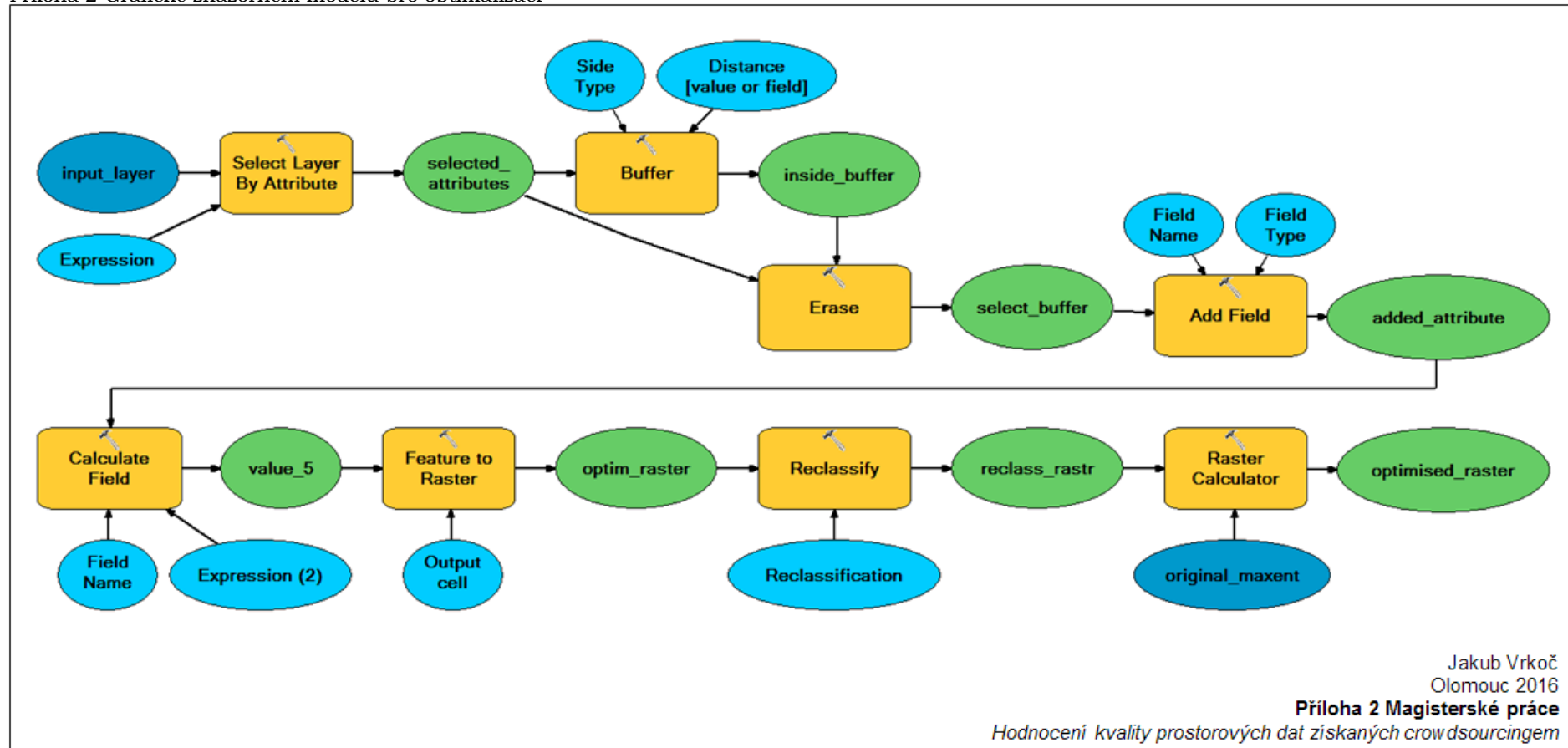


Jakub Vrkoč
Olomouc 2016

Příloha 1 Magisterské práce

Hodnocení kvality prostorových dat získaných crowdsourcingem

Příloha 2 Grafické znázornění modelu pro optimalizaci



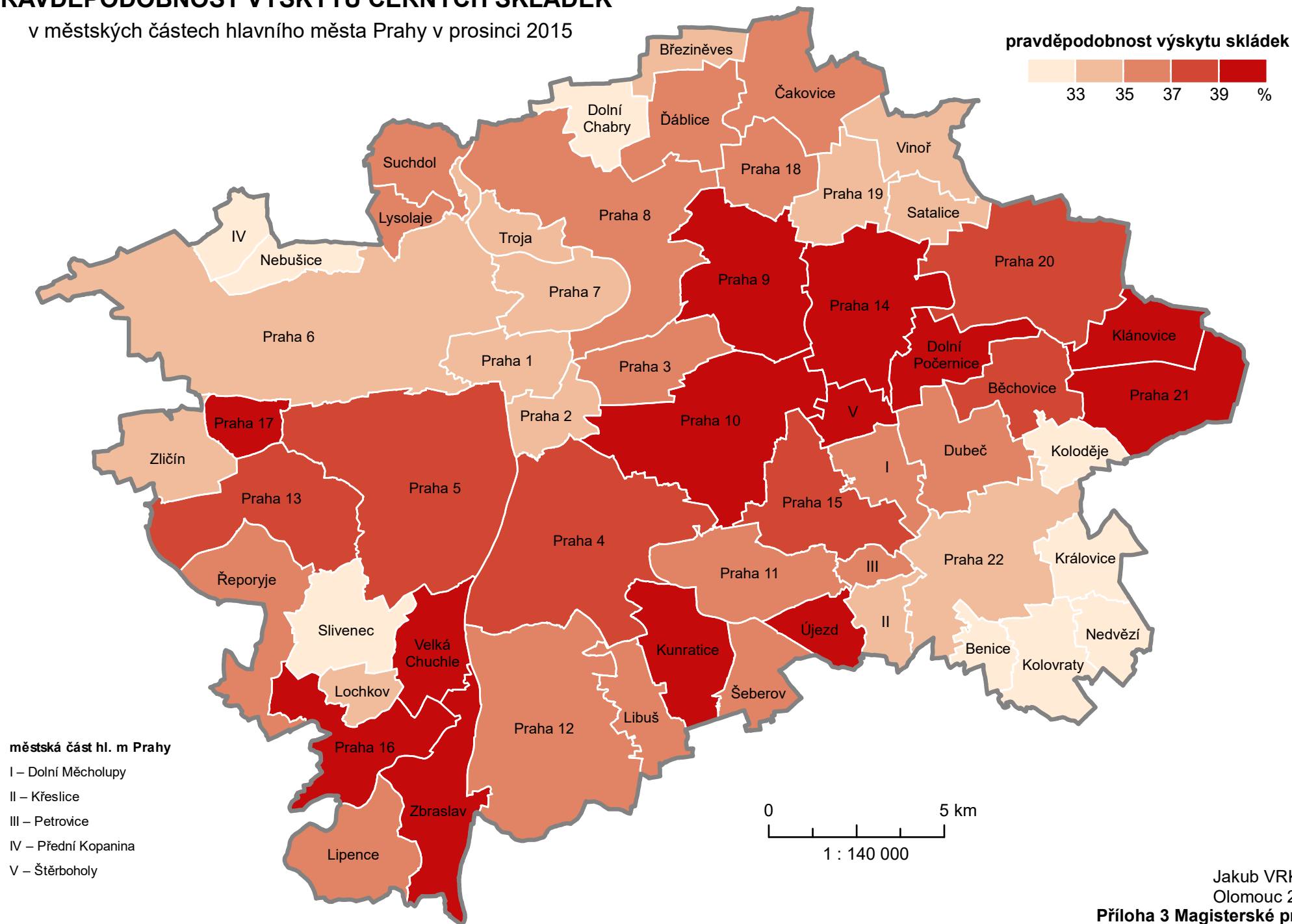
Jakub Vrkoč
Olomouc 2016

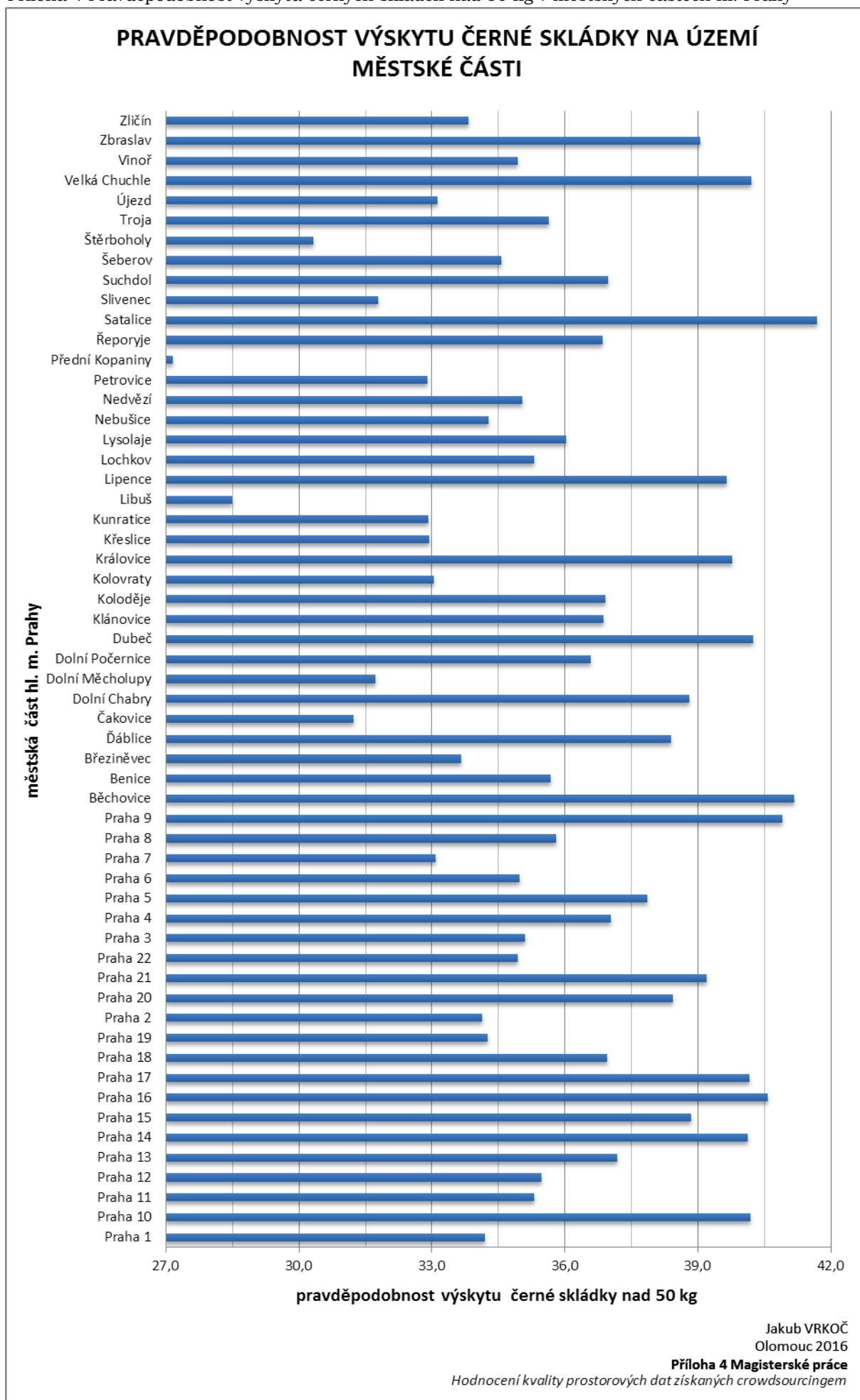
Příloha 2 Magisterské práce

Hodnocení kvality prostorových dat získaných crowdsourcigem

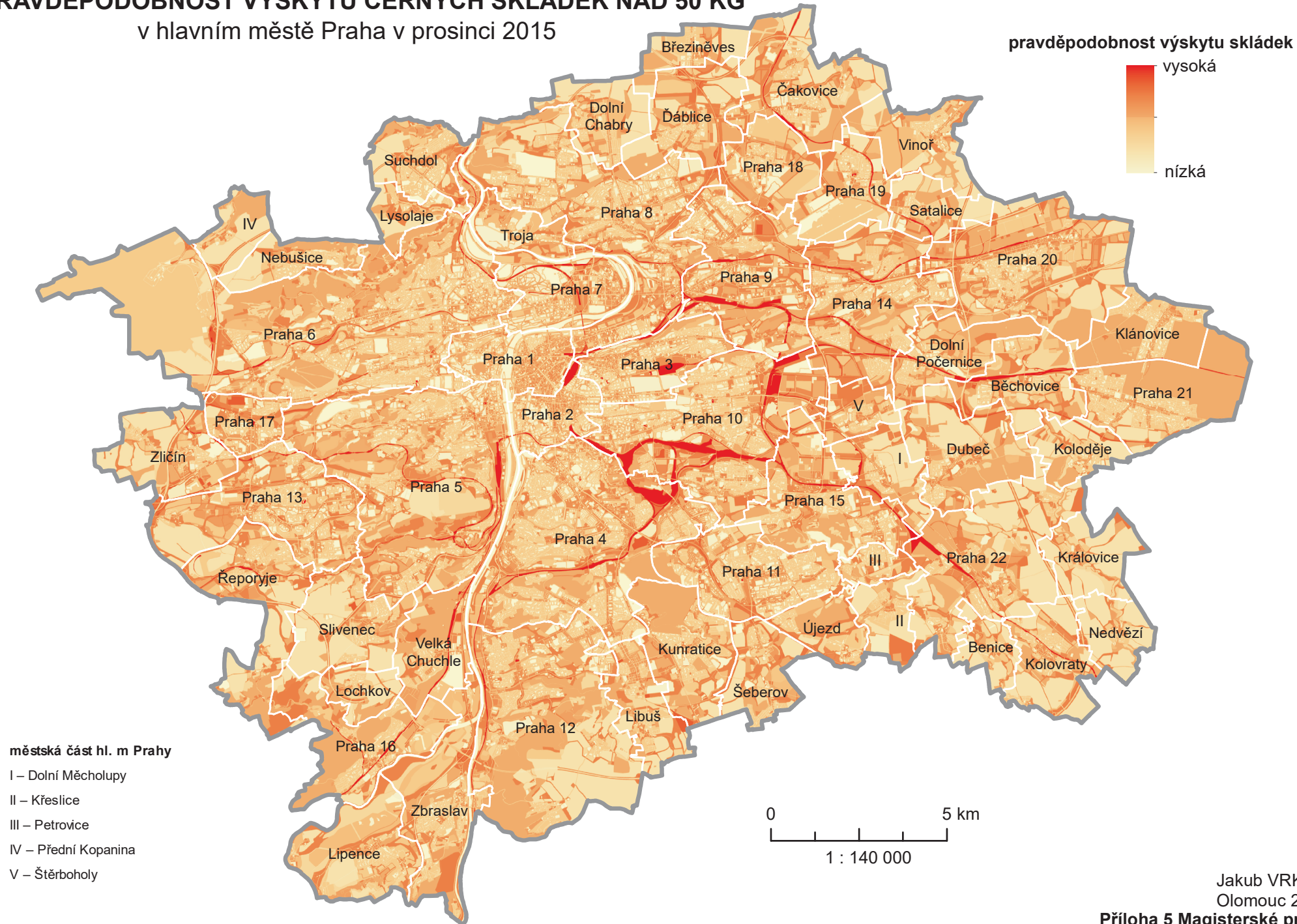
PRAVDĚPODOBNOST VÝSKYTU ČERNÝCH SKLÁDEK

v městských částech hlavního města Prahy v prosinci 2015





PRAVDĚPODOBNOST VÝSKYTU ČERNÝCH SKLÁDEK NAD 50 KG v hlavním městě Praha v prosinci 2015



Jakub VRKOČ
Olomouc 2016

Příloha 5 Magisterské práce

Hodnocení kvality prostorových dat získaných crowdsourcingem