

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

WEBOVÁ APLIKACE ZOBRAZUJÍCÍ POLOHU IP STANIC

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

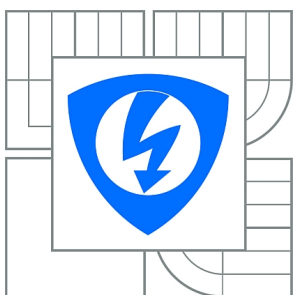
Bc. ZDENĚK MODRÁK

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ**

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

WEBOVÁ APLIKACE ZOBRAZUJÍCÍ POLOHU IP STANIC

WEB APPLICATION FOR GETTING LOCATION OF IP NODES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. ZDENĚK MODRÁK

VEDOUcí PRÁCE

SUPERVISOR

doc. Ing. DAN KOMOSNÝ, Ph.D.

BRNO 2015



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student: Bc. Zdeněk Modrák

ID: 134364

Ročník: 2

Akademický rok: 2014/2015

NÁZEV TÉMATU:

Webová aplikace zobrazující polohu IP stanic

POKYNY PRO VYPRACOVÁNÍ:

Vytvořte webový systém, který bude umožňovat zobrazení komplexních geografických informací pro zadané IP adresy včetně detekce přítomnosti proxy serveru. Webový systém realizujte v anglickém jazyce a umístěte jej na server Ústavu telekomunikací. Provádějte archivaci získaných informací. Z těchto informací zpracujte údaje týkající se dosahované přesnosti nalezení polohy. Dosažené výsledky publikujte na webových stránkách.

DOPORUČENÁ LITERATURA:

- [1] PUŽMANOVÁ, R. TCP/IP v kostce. 1. vyd. České Budějovice : Kopp, 2004. 607 s. ISBN 80-7232-236-2.
- [2] POESE, I., UHLIG, S., KAAFAR, M., DONNET, B., GUEYE, B. IP Geolocation Databases: Unreliable? ACM SIGCOMM Computer Communication Review. ACM, 2011.
- [3] GILMORE, J. Velká kniha PHP5 & MySQL: Kompendium znalostí pro začátečníky i profesionály. [překlad Miroslav Kučera; RNDr. Jan Pokorný]. 1. vyd. Brno: Zoner Press, 2005. 711 s. ISBN 80-86815-20-X.

Termín zadání: 9.2.2015

Termín odevzdání: 26.5.2015

Vedoucí práce: doc. Ing. Dan Komosný, Ph.D.

Konzultanti diplomové práce:

doc. Ing. Jiří Mišurec, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Práce se zabývá geolokací v síti internet. Postupně jsou rozebrány možnosti geolokace a dále se práce zabývá výhradně pasivními geolokačními metodami. Pod pasivní geolokační metody patří lokační databáze, které jsou v práci popsány, jak v teoretické rovině, tak i jsou použity v praktické části, kde je naprogramován komplexní webový systém s využitím placených a volně dostupných geolokačních databází. Dalším geolokačním zdrojem je databáze WHOIS. Jako doplňující data slouží naměřené zpoždění a detekce proxy. Data z placených geolokačních databází jsou dále zpracována a je vyhodnocena jejich přesnost.

KLÍČOVÁ SLOVA

Geolokace, IP geolokace, IP adresa, internet, WHOIS, lokační databáze

ABSTRACT

Thesis deal with geolocation in internet network. There are described possibilities of geolocation and thesis is mainly focused on passive geolocation methods. Under passive geolocation belongs location databases which there are described as in theoretical way as used in practical part of thesis. In practical part there is created complex system for geolocation in internet environment which used paid and free geolocation databases. Another used database is WHOIS. Data from paid databases is processed and accuracy of databases is evaluated.

KEYWORDS

Geolocation, IP geolocation, IP address, internet, WHOIS, location databases

MODRÁK, Zdeněk *WEBOVÁ APLIKACE ZOBRAZUJÍCÍ POLOHU IP STANIC*: diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2015. 57 s. Vedoucí práce byl doc. Ing. Dan Komosný, Ph.D.

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu doc. Ing. Danovi Komosnému, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno

.....

(podpis autora)



Faculty of Electrical Engineering
and Communication
Brno University of Technology
Purkynova 118, CZ-61200 Brno
Czech Republic
<http://www.six.feec.vutbr.cz>

PODĚKOVÁNÍ

Výzkum popsany v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno

.....

(podpis autora)



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



OBSAH

Úvod	9
1 Důvody k určení geografické polohy	10
2 Metody geolokace	13
2.1 Geolokace v sítích IP	14
2.1.1 Problematika adresace v sítích IP	15
2.1.2 Aktivní geolokační metody	16
2.1.3 Pasivní geolokační metody	16
2.2 Lokační databáze	18
2.2.1 Struktura geolokačních databází	18
2.2.2 Přístup k informacím v geolokačních databázích	19
2.2.3 Obecné informace o použitých geolokačních databázích	20
3 Zhotovený systém pro geolokaci stanic v síti internet	22
3.1 Popis programu	24
3.1.1 Vstupní parametry	24
3.1.2 Převod ulice a města na souřadnice	25
3.1.3 Výpočet odchylky	25
3.1.4 Získání dat z geolokačních databází	26
3.1.5 Grafická prezentace výsledků	36
3.1.6 Ukládání a zobrazení naměřených výsledků	37
4 Porovnání výsledků placených geolokačních databází	39
4.1 Zpracování výsledků geolokačních databází	40
4.1.1 Průměrná odchylka databází	40
4.1.2 Procentuální rozložení naměřených vzdáleností	41
4.1.3 Vyhodnocení efektivity lokace	44
4.1.4 Přesnost lokačních databází na základě správních celků	44
4.1.5 Zhodnocení výsledků	49
5 Závěr	52
Literatura	54
Seznam symbolů, veličin a zkratk	56
A Obsah příloženého CD	57

SEZNAM OBRÁZKŮ

1.1	Detekce polohy internetovým vyhledávačem Google.	12
2.1	Ukázka využití geolokace v mobilních telefonech.	13
2.2	NAT v síti internet.	15
3.1	Ukázka zadávání vstupních parametrů.	24
3.2	Vývojový diagram vytvořené aplikace.	35
3.3	Ukázka výsledků v tabulce.	36
3.4	Ukázka výsledků v mapě.	36
3.5	Výpis naměřených výsledků.	38
3.6	Zobrazení historie výsledků do mapy.	38
4.1	Poloha lokalizovaných IP adres na mapě.	39
4.2	Statistika odchylek lokalizovaných IP adres.	42
4.3	Chyba lokace v zobrazení po jednotlivých kategoriích.	43
4.4	Poměr lokalizace.	45
4.5	Kumulativní distribuční funkce pro lokační databáze.	46
4.6	Graf úspěšnosti detekce na základě správních celků.	50

SEZNAM TABULEK

3.1	Seznam použitých placených databází.	23
3.2	Seznam použitých volně dostupných databází.	23
4.1	Průměrná naměřená odchylka.	40
4.2	Odchylka pro nelokalizované IP adresy databází SkyHook Wireless.	41
4.3	Lokalizované IP adresy s chybou do 50 km.	42
4.4	Odchylka v procentech pro chybu mezi 100 až 250 km.	43
4.5	Úspěšnost detekce města nebo obce.	48
4.6	Úspěšnost detekce kraje.	49
4.7	Úspěšnost detekce na úrovni státu.	49
4.8	Bodové hodnocení výsledků.	50
4.9	Bodový zisk ve skupině procentuální rozložení naměřených odchylek.	51
4.10	Bodový zisk ve skupině rozpoznání správních celků.	51
4.11	Bodový zisk celkem i s body z průměrné naměřené odchylky.	51

ÚVOD

Internet je nedílnou součástí našich životů. Na internet se připojujeme odkudkoli. Z počítače, tabletu nebo mobilu, ať jsme doma nebo na cestách. S masivním rozmachem internetu souvisí i bezpečnost zařízení v síti. Jeden z pohledů na bezpečnost sítě může být geografická poloha útočníka. Metody geolokace jsou použity pro cílení obsahu jako je automatické nastavení města při vyhledání spojů MHD, zobrazování kulturních akcí ve městě. V internetu se používá především IP (Internet Protocol) geolokace. Tato metoda pracuje pouze s veřejnou IP adresou, která je v internetu unikátní pro každou stanici nebo skupinu stanic, obvykle sídlící na jednom místě. Existují metody pasivní a aktivní. Z nich jsou nejvíce používané právě pasivní metody. Pracuje se s databázemi čítající od několik stovek tisíc lokalizovaných IP adres, až po několik desítek milionů. Pro lokalizovanou adresu existuje záznam v databázi, kde mimo údaje o poloze stanice je obvykle obsaženo různé množství doplňujících informací. Existují volně dostupné databáze, které obvykle ve srovnání s placenými databázemi obsahují méně doplňujících informací. Nedosahují takové přesnosti jako placené databáze nebo nemají tak velký rozsah lokalizovaných IP adres.

Druhá kapitola popisuje možnosti pasivních geolokačních metod a jak je možné získávat geolokační data. Dále jsou zde rozebrány možnosti geolokačních databází, kterým se diplomová práce věnuje nejpodrobněji. Rozebrány jsou možnosti přístupu k datům. Každá databáze nabízí data v odlišném formátu a rozsah získaných informací je velmi rozdílný. Diplomová práce se bude zabývat placenými databázemi. Bude se jednat jak o online dostupné databáze, tak i databáze určené pro offline použití. To znamená, že offline databáze nepracuje s dalším vzdáleným serverem, ale vše je uloženo lokálně.

V praktické části diplomové práce je velmi podrobně popsána vytvořená aplikace v programovacím jazyku PHP (Hypertext Preprocessor), která zobrazuje přibližnou polohu IP stanice v síti. V jednotlivých podkapitolách je podrobně popsána funkce jednotlivých bloků aplikace. Pracuje se s několika placenými geolokačními databázemi a každá má jiný zdroj dat. Získaná data jsou prezentována formou tabulek, tak i vizualizací do mapového podkladu tak, aby výsledky bylo co možná nejnázorněji podány uživateli.

V poslední kapitole je popsána metoda testování a vyhodnocení přesnosti geolokačních databází tak, aby bylo možné zvolit databázi, která je schopná provést v praxi nejpřesnější geolokaci uživatele na základě IP adresy. Vyhodnocení výsledku je zvoleno tak, aby co nejvíce odpovídalo reálnému nasazení na serverech v síti internet.

Společně s touto prací pracoval na geolokačních metodách v síti internet také pan Janoušek. Jeho práce se zabývá volně dostupnými geolokačními databázemi.

1 DŮVODY K URČENÍ GEOGRAFICKÉ POLOHY

Proces sledování a určení geografické polohy, také často nazývaný jako geolokace, je v dnešním internetu nezbytný bezpečnostní prvek. Internet funguje jako nezávislá anonymní síť propojených počítačů s minimální kontrolou, ke které má neomezený přístup téměř kdokoli. S tím souvisí i otázka bezpečnosti. Po provedení kybernetického útoku bychom měli disponovat nástroji, které jsou schopné vystopovat útočníka zpět po případě část jeho sítě izolovat. Díky geolokaci jsme schopni odhalit útok dříve než k němu vůbec dojde. Správným nastavením bezpečnostních protokolů může být omezen přístup do sítě na základě polohy přistupujícího uživatele. Na základě již odhalených útoků a porovnáním polohy, nově navazující se spojení z oblasti častých útoků lze ihned označit za podezřelé a lze toto spojení podrobit hlubší analýze.

Chceme-li zjistit geografickou polohu zařízení v sítích internetu, typicky se jedná o počítač, tablet nebo mobilní telefon, používá se technika s názvem IP geolokace. Pracuje se pouze s veřejnou IP adresou zařízení připojeného do sítě internetu. Veřejná adresa je unikátní identifikátor komunikujícího zařízení nebo stále častěji skupiny komunikujícího zařízení, která může obsahovat jednotky až tisíce zařízení. Pomocí této techniky jsme schopni zaměřit zeměpisné souřadnice, ale také lze detekovat stát, město, ulici a k těmto datům můžeme zjistit i další informace jako je PSČ, peněžní měna.

Geolokace může být použita i v případě ochrany finančních prostředků v bankách. Banky mohou omezit přihlašování k internetovému bankovníctví ze zemí ve kterých jejich pobočky nepůsobí nebo platby kartou mohou být z cizích zemí omezeny nebo úplně zakázány. Lze také ověřit věrohodnost emailu porovnáním místa vzniku hlavičky emailu a místa vzniku těla emailu. [1]

S rozšiřováním sítě internetu je zakládána spousta webů, které fungují pouze elektronicky a pro jejich autory je reklama na webu často jedinou formou výděлку. Internetoví inzerenti vyžadují co nejpřesnější zacílení reklamy na uživatele. Jeden z faktorů, který ovlivňuje přesnost zacílení reklamy a s tím související potenciální možnost kliknutí uživatele na reklamu, je aktuální poloha uživatele. Inzerent chce propagovat své produkty pouze ve svém regionu nebo velké řetězce chtějí inzerovat jejich akční nabídky v nejbližším nákupním centru. Je mnohem výhodnější pro inzerenta nabízet produkty dostupné v dané oblasti než nabízet produkty dostupné po celém světě.

S rostoucí oblibou streamovacích služeb nebo jiných služeb náročných na kapacitu linky nebo na odezvu serveru, je vhodné nabídnout uživateli obsah z nejbližšího dostupného webového serveru z důvodu rozložení zátěže v síti poskytovatele ob-

sahu. S poskytováním obsahu souvisí i ochrana autorských práv. Autorská práva jsou nakupována pro jednotlivé země nebo části světa. Obyvatelé z USA si tak často nemohou pustit obsah poskytovaný v Evropě a naopak. S výhodou tuto službu také využívají národní televizní stanice, které omezují streamování obsahu pouze na zemi působnosti. [2]

S rozvojem VoIP (Voice over Internet Protocol) telefonie se objevila komplikace při realizaci tísňových hovorů, kdy nebyla známá ani přibližná poloha volajícího, docházelo k nesprávnému směrování na nesprávný dispečink. Následným přepojováním dochází ke zdržení a k možnému ohrožení života. Pokud by volající nebyl schopný popsat svoji aktuální polohu může automatické určení polohy zachránit život. [3]

Dalším možným využitím známé polohy uživatele je možnost, návštěvníkovi stránek nabídnout jejich obsah automaticky v jeho rodné řeči. Nejvíce známým cílím informací na základě polohy může být uživatelům známé zobrazování aktuálních informací, ať se jedná o počasí, tabule odjezdů vlaků z nejbližší vlakové stanice a to automaticky a bez zásahu uživatele.

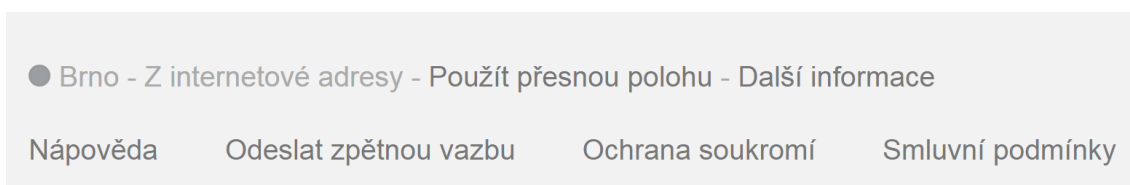
Je-li známá aktuální poloha v systému iOS, nemusí se jednat přímo o přesnou polohu ze systému GPS (Global Positioning System). Systému stačí přibližná poloha získána například z údajů o dostupných Wi-Fi sítích, systém pak automaticky nabídne zajímavé aplikace nebo webové stránky související s aktuálním místem. V Brně se jedná například o okolí hlavního nádraží. Jakmile se nacházíte v jeho blízkosti, systém velmi nenásilně nabízí aplikaci Český drah, kde je možné vyhledat vlakové spojení.

Mobilní telefon je schopný neustále sledovat polohu zařízení. Jakmile opustíte definovaný prostor, například kancelář, telefon akci zaznamená a může zaslat upozornění na definovaný úkol. S výhodou, lze tuto funkci využít tam, kde časově definované připomínky přestávají stačit.

Na sociálních sítích, jako je Facebook nebo Twitter, je sdílení polohy ve výchozím nastavení zapnuto. Není tedy problém zjistit, kde se uživatel, se kterým jsme v kontaktu, nachází. Na Facebooku můžeme přidat svoji polohu k příspěvku, který se chystáme zveřejnit. Nebo ve Facebook Messengeru je ke každé odeslané zprávě odeslané z mobilního zařízení přiložena poloha.

Na internetu vznikla spousta geolokačních her. Smysl těchto her je nalákat uživatele ven a jako odměnu jim za daný počet takzvaných check-inu rozdávat různá ocenění. Na tomto principu funguje například Swarm. K vlastnímu check-inu lze přidávat komentáře, fotografie a hodnocení, které může pomoci rozhodnout, zda místo navštívit nebo jít raději jinam. Další geolokační hrou je ve světě velmi známý Geocaching. Jedná se v principu o velmi jednoduchý systém. Na nějakém zajímavém místě uživatel schová různě velikou krabičku (keš), kterou zaměří a souřadnice odešle na server. Úkolem ostatních uživatelů je najít schovanou keš a zapsat se do log booku.

S čím dál tím větší četností využívání IP geolokace se lze setkat u vyhledávačů, kteří se snaží odhadnout aktuální polohu počítače tak, aby byly schopni nabídnout uživateli co nejrelevantnější informace. V aktuálním nastavení má poloha uživatele velmi velkou váhu na pořadí výsledků. Při hledání řetězce rychlého občerstvení je na prvním místě oficiální stránka, ovšem na druhém místě je již seznam provozoven nacházející se v daném městě nebo alespoň poblíž detekovaného města. Polohu ve vyhledávači Google lze nastavit ručně. Pokud poloha není nastavena, vyhledávač detekuje polohu z IP adresy, pomocí funkce historie polohy nebo podle nedávno vyhledávaných míst.



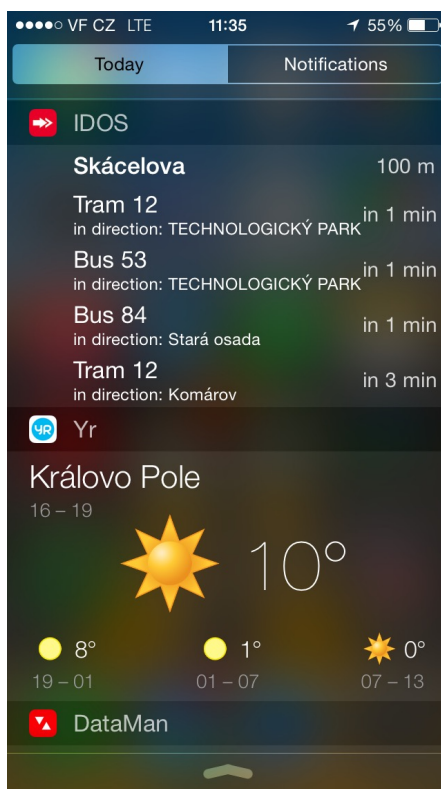
Obr. 1.1: Detekce polohy internetovým vyhledávačem Google.

2 METODY GEOLOKACE

Obecně lze říci, že existuje několik metod geolokace, které využívají různých prostředků a služeb pro určení reálné polohy uživatele nebo jeho stanice. Různé metody pracují s rozdílnou přesností. Pro úplnost jsou zde uvedeny metody geolokace pomocí BTS (Base Transceiver Station) a GPS, ale dále s nimi tato diplomová práce nebude pracovat. Při určování polohy pomocí BTS se obvykle pracuje se známou polohou BTS, se kterou mobilní telefon právě komunikuje. Přesnost je dána poloměrem signálu dané BTS. Ve městech může přesnost dosáhnout až 150 metrů, na venkově nebo v řídko obydlených oblastech se může jednat až o jednotky kilometrů. [5]

Určení polohy pomocí GPS

Určování polohy pomocí GPS se v poslední době velmi rozšiřuje. GPS čip obsahuje každý chytrý telefon, oblast působnosti se rozšiřuje i do moderních fotoaparátů a chytrých hodinek. GPS je družicový polohový systém provozovaný ministerstvem obrany Spojených států amerických. Přesnost systému GPS se pro civilní uživatele pohybuje v jednotkách metrů, avšak s použitím dalších systémů, lze přesnost zvětšit až na jednotky centimetrů. [6]



Obr. 2.1: Ukázka využití geolokace v mobilních telefonech.

Přibližné určení polohy pomocí Wi-Fi sítě

Google na svém mapovém portálu provozuje službu StreetView, která nabízí panoramatické snímky z mnoha zemí a měst. Panoramatické snímky jsou pořizované automobily a nepřístupná místa jsou snímány tříkolkami nebo sněžnými skútry. Mimo zachytávání panoramatických snímků jsou sbírány i informace o přítomných Wi-Fi sítích. Údaje o Wi-Fi sítích obsahují jejich MAC (Media Access Control) adresu, SSID (Service Set Identifier) a polohu. Jelikož tyto informace velice rychle zastarávají a dnes je doba chytrých telefonů, využívá se jejich všudypřítomnost k zdokonalení a udržení aktuálních informací v databázi. Chytrý telefon neustále skenuje Wi-Fi pásmo, rádiovou část mobilní sítě, pokud jsou dostupné polohové služby, tak využívá i tyto data. Sesbírané informace telefon anonymně odesílá na servery výrobce operačního systému. Vlastní databáze má Google, Apple i Microsoft. Tyto informace jsou naopak využívány uživateli mobilních telefonů pro rychlé zaměření jejich polohy, když je to vyžadováno. [7]

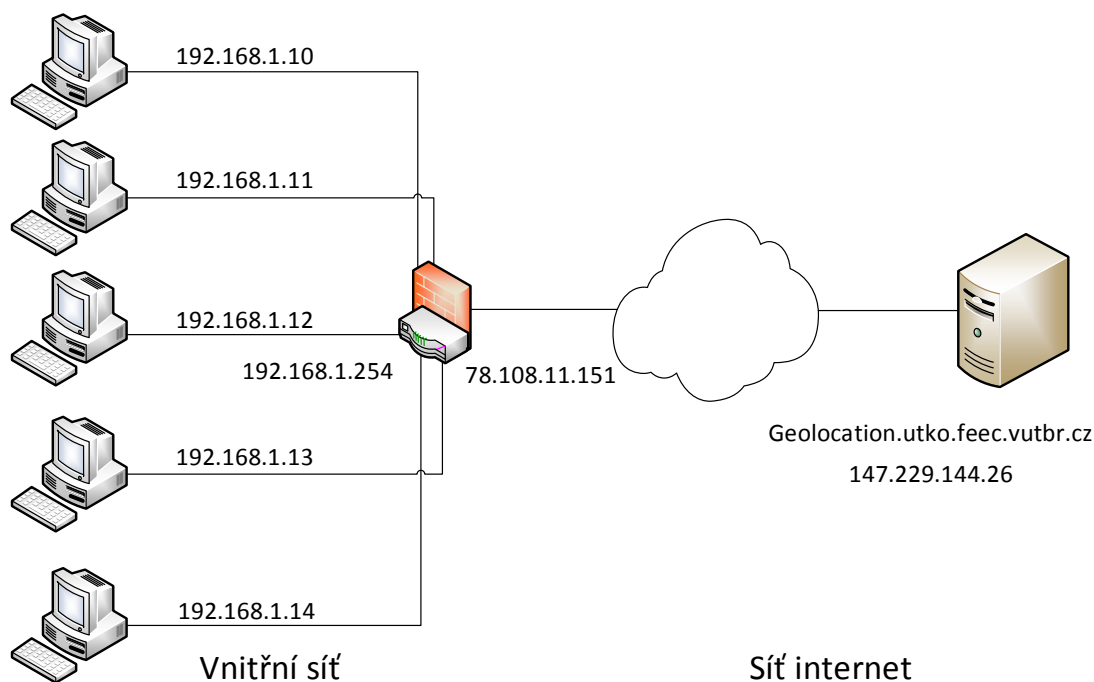
2.1 Geolokace v sítích IP

Existuje řada problémů ve vyhledání správné geografické polohy stanice v internetu. Z pohledu určování polohy v síti IP je nejzřetelnější, že neexistuje žádný internetový protokol, který by poskytoval aktuální polohu každého zařízení v síti. Pravdou je, že DNS (Domain Name System) může obsahovat informaci o poloze zařízení. Další komplikací je, že zařízení přistupující k síti nejsou typicky vybavena identifikátorem jejich polohy. S nástupem technologie GPS do mobilních telefonů, tabletů a postupným rozšířením působnosti GPS se situace pomalu mění. Pak ale přesnou polohu těchto zařízení chceme uchovat v soukromí. Geolokační metody rozdělujeme podle způsobu zjišťování do dvou skupin. Na metody aktivní a pasivní. Aktivní metoda využívá měření zpoždění (latence) přenosu dat sítě od odesílatele s aktuálně neznámou polohou k příjemci s polohou známou. A následně jsou naměřená data přepočítána na aktuální polohu stanice. Pasivní metody, na rozdíl od aktivních, neprovádí žádné měření, ale podle parametrů stanice je poloha vyhledána v lokačních databázích.

Stanice v síti IP nezná svoji polohu ani ji sama automaticky nevyhledává. Polohu stanice vyžaduje server, se kterým stanice navazuje komunikaci. Získaná data o poloze server zpracovává a může pozměnit charakter nabízených služeb.

2.1.1 Problematika adresace v sítích IP

Ideální situací pro geolokaci by byla, kdyby každé zařízení v síti mělo vlastní veřejnou IP adresu, aby bylo možné přesně trasovat cestu paketu od zdroje k cíli. Z důvodu úspory veřejných IP adres, kterých je v IPv4 znatelný nedostatek jsou zaváděny rozsáhlé privátní sítě, které pak v internetu vystupují pod jednou veřejnou IP adresou. Vysoký počet privátních sítí je v domácnostech, ale nejsou výjimkou ani rozsáhlé metropolitní sítě, které pokrývají rozsáhlá území měst a okolních obcí. Zařízení uživatelů se nacházejí za NAT (Network Address Translation), příklad je znázorněn na obrázku 2.2. Jedná se o postup, který umožňuje připojit více zařízení s privátními adresami do internetu, ve kterém se použije pouze jedna veřejná adresa pro všechna zařízení. Společně s NAT se aplikují také pravidla pro překlad portů PAT (Port Address Translation).



Obr. 2.2: NAT v síti internet.

Z hlediska geolokace zde vzniká problém, jak správně určit polohu zařízení. Ze sítě internet, kde se nachází geolokační server, který sloužil pro aktivní geolokační metody, by privátní adresu bez aplikování specifických pravidel na hraničním routeru s NAT nikdy nemohl zaměřit. V situaci, kde se za NAT nachází pouze několik zařízení, které se nachází v jedné budově, zde není problém s geolokací. V případě, že se za NAT nachází rozsáhlá metropolitní síť, je geolokace velmi obtížná, respektive, zde se otvírá veliký prostor pro vznik nepřesností a chyb. Z výsledků, které jsou uvedeny

dále vyplývá, že jako výsledek geolokace je ve velkém množství výsledků uvedena adresa provozovny poskytovatele připojení. S reálnou polohou uživatele v těchto případech nemá detekovaná poloha mnohdy nic společného. Také je důležité zmínit, že neexistuje mechanismus, který by umožnil zohlednit reálnou polohu uživatele.

2.1.2 Aktivní geolokační metody

U aktivních geolokačních metod odhadujeme polohu měřením parametrů datového provozu v internetové síti. Nejčastěji je měřeno zpoždění a je zjišťována trasa přes mezilehlé uzly. Tato diplomová práce se zabývá podrobně pouze pasivními metodami. Budou zde uvedeny pouze některé nástroje využívající aktivní metody k určení polohy. Těmi úplně nejzákladnějšími nástroji je nástroj Ping a Traceroute. Zpoždění je měřeno v obou směrech, protože trasa od zdroje k cíli může být odlišná než trasa v opačném směru od cíli ke zdroji.

Pokročilejší metody jsou následující [3]:

- Octant,
- NetGeo,
- IP2Geo - GeoPing,
- Constraint Based Geolocation (CBG),
- Topology Based Geolocation (TBG).

2.1.3 Pasivní geolokační metody

Pasivní geolokační techniky jsou jednodušší než aktivní, ale jejich hlavní nevýhodou je, že pro své správné fungování vyžadují správné údaje v geolokačních databázích. Aktualizace údajů probíhá automaticky, ale přesto je časově náročná a je zde velký prostor pro vznik chyb. Vytvořit velmi přesné algoritmy je velmi finančně náročné. Z toho důvodu je většina databází komerčních. Obvykle vznikají dva druhy nepřesností a to chybou ve vstupních datech a chyba neohlášenou změnou přidělení IP adresy. Neprobíhá-li aktualizace dat, tak databáze neprovádí žádná další měření, pouze se využívají data uložená v databázích.

Využití informací o přidělování IP adres organizací IANA

Řízení přidělování IP adres v celosvětovém měřítku, správu kořenových DNS serverů a další aktivity spojené s fungováním internetu má na starosti IANA (Internet Assigned Number Authority), která spadá pod neziskovou organizaci ICANN (Internet Corporation for Assigned Names and Numbers) se sídlem v americké Kalifornii [8].

IANA se dále dělí na několik koordinačních středisek na takzvané RIR (Regional Internet Registry). Každý registr má na starosti jinou část světa.

Pro Evropský kontinent, střední východ a části střední Asie vznikla v roce 1992 mezinárodní nevládní organizace RIPE NCC (Réseaux IP Européens Network Coordination Centre), která je součástí RIR a má na starost přidělování IP adres spadajícím do LIR (Local Internet Registry). LIR jsou poskytovatelé internetu nebo velké organizace, které si staví svá vlastní datacentra nebo mají své linky například mezi pobočkami společností.

IANA přiděluje bloky IP adres, typicky jsou to adresy s maskou sítě /8. Z dokumentů lze pak vyčíst, že s počátkem rychlého rozmachu internetu mezi roky 1992 – 1995 byly bloky /8 přidělovány i jednotlivým společnostem. První bloky pro RIPE NCC byly přiděleny v roce 1993, další pak v roce 1997. [9]

RIPE NCC provozuje rozsáhlé volně přístupné databáze, kde je možné dohledat, kdy a komu je právě přidělený jaký rozsah IP adres. Na základě těchto údajů je možné provést prvotní plnění geolokačních databází s relativně vysokou přesností. Míra přesnosti závisí na velikosti společnosti a její oblasti působnosti. Telekomunikační společnosti mají oblast působnosti na úrovni státu a naopak datacentra, která pokrývají velkou škálu služeb, se nacházejí obvykle v jedné či více sousedících budovách. Přesnost údajů obsažených v databázích jednotlivých RIR v oblasti geolokace dosahuje přesnosti minimálně na úrovni státu. U některých údajů lze přesnost zvýšit až na ulici a číslo popisné.

Údaje získané z databází IANA respektive z RIR je nutné neustále aktualizovat. Přidělené bloky IP adres i na úrovni organizace IANA se mohou změnit a to kdykoli, kdy to situace bude vyžadovat. V důsledku nedostatku IP adres se bloky IP adres přelívají z jednoho LIR do druhého a dochází k optimalizaci přidělených IP adres. Také veřejné IP adresy na úrovni ISP se mohou měnit. ISP obvykle funguje na dvou modelech práce s veřejnými IP adresami. První model funguje tak, že uživatel je za NATem poskytovatele se všemi ostatními uživateli. V druhém modelu také figuruje NAT, ale v mnohem menším měřítku a to přímo na uživatelově routeru. Router má přidělenou IP adresu, pod kterou uživatelé přistupují k internetu. Tato veřejná IP adresa se mnohdy velmi často mění z rozsahu IP adres přidělenému poskytovateli.

Dělení organizace RIR je následovné:

- Africké síťové informační centrum (AfriNIC),
- Americký Registr pro Internet Numbers (ARIN),
- Asijsko-Pacifické síťové informační centrum (APNIC),
- Evropské síťové koordinační centrum (RIPE NCC),
- Latinsko-Americké a Karibské Internetové adresy registrů (LACNIC).

2.2 Lokační databáze

Nejvíce efektivní způsob z hlediska implementace, jak zjistit polohu přístupujícího návštěvníka na webový server, je využití některé z mnoha geolokačních databází. Získané informace lze využít k automatickému nastavení jazyka, poloha může sloužit jako jeden z parametrů k zobrazování reklamy anebo k dalším účelům popsáných v první kapitole. Obecně lze geolokační databáze rozdělit do dvou skupin a to volně dostupné a placené databáze. Dále lze databáze rozdělit na dostupné online a nebo offline. Na problematiku, zda využít offline a nebo online databázi, lze nahlížet ze dvou směrů. Online databáze proti offline databázi by měli obsahovat aktuálnější informace, protože offline databáze vycházejí v určitých periodách. Tím výčet výhod pro online databáze končí. Kritickým hlediskem je zde dostupnost takové databáze, která nikdy nedosáhne 100 %. Programátor by si tedy měl uvědomit, zda je určení polohy pro něj prioritní. Rychlost zpracování dotazu online databází může v čase kolísat od několika milisekund do několika sekund. Dalším omezením, které musí být zohledněno, je počet dotazů za jednotku času, které můžeme na geolokační server zaslat. U volně dostupných databází často dochází k blokování služby po překročení limitu dotazů. U placených databází s počtem dotazů roste měsíční paušál za využívání služby. Offline databáze lze automaticky aktualizovat pomocí periodicky zasílaných dotazů a také offline databáze bude mít stejnou dostupnost jako hostovaná webová stránka.

2.2.1 Struktura geolokačních databází

U placených databází není zcela dostupná dokumentace o tom, jak databáze pracují a jaká je jejich struktura, pouze některé databáze nabízejí strohé informace o tom, jak ukládají získané informace. Nedostupnost informací je pochopitelný krok, který zamezuje únikům informací, útokům na databáze a může být i konkurenční výhodou. Jejich API se snaží pracovat formou „Black boxu“, kde se uživatel nemusí starat o to, co tom co děje v „backendu“ a zajímá se pouze o vstupní a výstupní data.

Databáze DB-IP.com nabízí instalačního průvodce, ze kterého lze získat informace o struktuře databáze. Struktura není nijak komplikovaná, jak by mohlo z fungování databází vyplývat. V instalačním průvodci je vytvořena jedna databáze s jedinou tabulkou, ve které se nacházejí veškeré informace. Skript vytvářející strukturu tabulky je uveden ve výpisu kódu 2.1. DB-IP.com nabízí více druhů databází s různým rozsahem obsahu. Nejmenší tabulka obsahuje pět sloupců s daty, naopak nejobsáhlejší obsahuje třináct sloupců. Následuje MySQL skript pro vytvoření nejobsáhlejší tabulky. [10]

```

1 CREATE TABLE 'dbip_lookup' (
2   'addr_type' enum('ipv4','ipv6') NOT NULL,
3   'ip_start' varbinary(16) NOT NULL,
4   'ip_end' varbinary(16) NOT NULL,
5   'country' char(2) NOT NULL,
6   'stateprov' varchar(80) NOT NULL,
7   'city' varchar(80) NOT NULL,
8   'latitude' float NOT NULL,
9   'longitude' float NOT NULL,
10  'timezone_offset' float NOT NULL,
11  'timezone_name' varchar(64) NOT NULL,
12  'isp_name' varchar(128) NOT NULL,
13  'connection_type' enum('dialup','isdn','cable',
14                        'dsl','fttx','wireless') DEFAULT NULL,
15  'organization_name' varchar(128) NOT NULL,
16  PRIMARY KEY ('ip_start'));

```

Výpis kódu 2.1: Skript vytvářející strukturu tabulky v databázi DB-IP.

Databáze pracuje s IPv4 i s IPv6 adresami, které jsou definovány jako blok IP adres s určitým rozsahem. Počátek bloku je definován v buňce `ip_start` konec je definován v buňce `ip_end`. Tento přístup vede ke značné efektivitě při ukládání výsledků v databázi a vede k rychlému vyhledávání ve výsledcích. Návrh tabulky není ukázkovým příkladem, jak by se měla vytvářet optimální databáze. V tabulce bude docházet k mírné duplicitě výsledků a to především v polích `stateprov`, `city`, `timezone_name` a `organization_name`. Výhodnější by bylo vytvořit pro každou zmíněnou položku zvláštní tabulku s názvy a v tabulce uvedené výše, by byly uvedeny pouze ID odkazy, tím by se zamezilo duplicitám a dosáhlo by se větší úspory místa a rychlejšímu načítání dat.

2.2.2 Přístup k informacím v geolokačních databázích

Z geolokačních databází lze získat široký rozsah informací. Nejobsáhlejší databáze obsahují až 17 parametrů o zjišťované IP adrese. Mezi ty základní lze zařadit Country Code, které můžeme získat například z online databáze Geobytes příkazem `$response['iso2']`. Výstupem je kód země. Je-li přistupující uživatel z České republiky, výstupní proměnná obsahuje CZ. Mezi běžně dostupná data patří název města `$response['city']`, zeměpisná šířka `$response['latitude']` a zeměpisná délka `$response['longitude']`. Takováto data jsou dostatečná k dostatečně přesnému určení polohy. Obsáhlejší databáze obsahují název poskytovatele připojení,

nebo kód měny, který může být použit u mezinárodních eshopů. Mezi užitečné údaje může také patřit detekce proxy serveru.

Tato diplomová práce se bude zabývat hlavně vyhodnocením přesnosti a pravdivosti údajů obsažených v placených databázích. Výhodou těchto databází je, že obsahují údaje z celého světa a ne pouze data z větších zemí. Každý vydavatel tvrdí, že poskytují nejpřesnější možné údaje, ale rozptyl poskytovaných údajů je značný. Není problém nalézt IP adresu, kde je rozptyl výsledků v řádech tisíců kilometrů.

Placené databáze lze použít v licenci trial nebo demo. Někteří tvůrci databází bohužel tuto možnost nenabízejí. Databáze v tomto režimu neobsahují kompletní data. Chybějící údaje jsou například o poskytovateli připojení nebo pokud se jedná o online dostupnou databázi, tak je omezen počet údajů za hodinu.

2.2.3 Obecné informace o použitých geolokačních databázích

Skyhook Wireless

Skyhook poskytuje online přístupnou geolokační databázi bez omezení po registraci pouze pro IPv4 ve více než 200 zemích. Směrovací čísla a demografické údaje jsou dostupná pouze v USA a v Austrálii. Další země budou brzy následovat. Databáze dokáže poskytnout název země v krátkém i v dlouhém formátu, stát, poštovní směrovací číslo, zeměpisnou délku a zeměpisnou šířku.[11]

The DB-IP Database

Obsahuje geolokační záznamy jak pro IPv4, tak nově i pro IPv6 adresy. S téměř sedmi miliony lokalizovanými IP adresami patří k větším hráčům na trhu. Měsíčně dojde k úpravě až u 900 000 záznamů. Databáze je v nejnižší verzi dostupná zdarma, avšak jsou zde jistá omezení. Počet dotazů je omezen na 2 000 dotazů za den a jsou poskytovány pouze nejzákladnější data jako je země, stát a město. Další komplikací je chybná interpretace respektive žádná implementace diakritiky.[12]

Geobytes

Geobytes nabízí produkt IP Address Locator. V zimním semestru Geobytes nabízel značně omezený geolokační systém. Prvních 20 dotazů za hodinu bylo zdarma, poté bylo nutné zakoupit kredit. V opačném případě byla služba zablokována. Během semestru byla služba přepracována na nový model a nyní je nabízena zcela zdarma. Geobytes nabízí širokou škálu výsledků o dané IP adrese. Mezi základní patří kód země, region, město a zeměpisná šířka se zeměpisnou délkou. Mezi doplňující patří měna, zda-li se jedná o proxy IP adresu, časový posun a populace státu. [13]

MaxMind GeoLite2

Databáze GeoLite2 je offline volně dostupnou verzí IP geolokační databáze od firmy MaxMind. Jako placená verze je nabízena GeoIP2. Obě databáze obsahují srovnatelná data jak pro IPv4 tak i pro IPv6. Ovšem placená verze by měla být přesnější. Lite verze je aktualizována jednou měsíčně. Pro zjednodušení práce je možné využít automatizovaných nástrojů pro aktualizaci dat v databázi přímo od firmy Maxmind a mít tak stále aktuální data. Databáze GeoIP2 patří s pokrytím 99,99 % adresního rozsahu k těm největším a nejpresnějším na trhu. Denně je vyhodnocováno a aktualizováno až 30 milionů IP adres. Databáze poskytuje záznamy o kontinentu, státu, městu a geografické poloze. Dalšími poskytovanými informacemi je poštovní směrovací číslo a časové pásmo. Placená verze může dále obsahovat poskytovatele připojení, doménové jméno, číslo autonomního systému, zda je připojení poskytnuté satelitem a detekci proxy. [14]

IP Info

Je placenou databází s 1 000 dotazy za den zdarma přes nešifrované spojení HTTP (Hypertext Transfer Protocol). Placená verze je dostupná v několika verzích podle počtu dotazů za den s podporou zašifrovaného přenosu HTTPS (Hypertext Transfer Protocol Secure). Nabízí také vyšší přesnost výsledků a detekci proxy. Obě verze podporují jak IPv4 tak i IPv6. Databáze obsahuje jak základní informace, tak i doplňující informace. Mezi zajímavé údaje patří hostname IP adresy, číslo autonomního systému a telefonní předvolba v oblasti. [15]

3 ZHOTOVENÝ SYSTÉM PRO GEOLOKACI STANIC V SÍTI INTERNET

Cílem webové aplikace je vytvořit komplexní web věnující se do maximální možné míry geolokaci na základě IP adresy. Mezi zjišťované parametry patří zejména geolokace podle dostupných geolokačních databází, měření zpoždění a detekce polohy na základě údajů obsažených v databázích WHOIS. Práce na projektu byly rozděleny na dvě části. V prvním kroku byl zhotoven web, který zajišťoval detekci pouze na základě IP adresy pomocí placených geolokačních databází. V druhém kroku byly přidány volně dostupné geolokační servery, na kterých pracoval kolega Janoušek. Za další bylo přidáno měření zpoždění k cílové stanici ze serveru umístěného na VUT v Brně na kterém spolupracoval kolega Bednář. Předposledními daty získávány externě jsou data z databáze WHOIS od pana kolegy Smrčky, která jsou prezentována společně s daty získanými z geolokačních databází ve stejné tabulce. Aplikace jako celek tedy používá devět na sobě nezávislých zdrojů, ze kterých lze odhadnout polohu zadané IP adresy. Jako poslední dodal svá data pan Gardian, který ve své bakalářské práci detekuje přítomnost proxy na základě IP adresy.

Pracuje-li se s online dostupnými lokačními databázemi, je přímo ve webovém prohlížeči možné zadat ve specifikaci API danou URL. Obvykle řetězec obsahuje identifikační klíč a zjišťovanou IP adresu. Po odeslání URL na server je vypsan přímo textový výsledek z databáze do okna prohlížeče. Obvykle se jedná o data ve formátu JSON (JavaScript Object Notation) nebo XML (eXtensible Markup Language). U offline databází lze využít nástrojů přímo od dodavatele, výstupem bude požadovaný řetězec dat. Z výhodou lze využít, že offline databáze jsou dostupné jako jeden soubor a nemusíme znát vnitřní strukturu. I aktualizace takové databáze bude velmi jednoduchá a lze ji velmi snadno zautomatizovat.

V prvním kroku bylo zhotoveno vyhledávání v placených databázích. Pro vyhledávání polohy je využito pět geolokačních databází z nich byly čtyři geolokační databáze dostupné online a jedna databáze byla dostupná off-line. Tato databáze je umístěna na stejném serveru, jako je umístěn vlastní program. Seznam použitých databází je uveden v tabulce 3.1:

V druhém kroku byly odstraněny známé chyby, byl přidán překlad doménových jmen na IP adresy a hlavně aplikace byla rozšířena o volně dostupné geolokační databáze za přispění pana Janouška. Seznam použitých volně dostupných databází je uveden v tabulce 3.2: Výsledky měření jsou zobrazeny na webu ve stejné tabulce společně s placenými lokačními databázemi a jsou společně zobrazeny v mapě. Takto prezentována jsou i data získána z databáze WHOIS. Samostatnou tabulku tvoří takzvané doplňkové informace. Jedná se o naměřené zpoždění v milisekun-

dách. V případě, že se měření nezdařilo, je vypsáno hlášení `not available`. Jako druhý údaj v tabulce s doplňkovými informacemi je zobrazené naměřené zpoždění ze serveru na VUT v Brně.

Tab. 3.1: Seznam použitých placených databází.

Název	Typ databáze
Skyhook Wireless	Online databáze
The DB-IP Database	Online databáze
Geobytes	Online databáze
Maxmind GeoLite2	Offline databáze
IP Info	Online databáze

Tab. 3.2: Seznam použitých volně dostupných databází.

Název	Typ databáze
FreeGeoIP	Online databáze
HostIP.info	Online databáze
IP2Location	Offline databáze

Vlastní aplikace funguje jako webová stránka na školním serveru Geolocation <<http://geolocation.utko.feec.vutbr.cz/xmodra00/>>. Realizace aplikace probíhala ve velmi rozšířeném programovacím jazyku PHP. Stránky byly doplněny o jazyk JavaScript, zajišťující pohodlnější ovládání a lépe vypadající vstup. Samotný vzhled stránek je generován jazykem HTML (HyperText Markup Language) a dále je upraven CSS (Cascading Style Sheets). Definice vzhledu jsou obsaženy v souboru `style.css`. Offline databáze Maxmind GeoLite2 obsahuje knihovnu pro PHP. Její součásti jsou nahrány na webovém serveru. Soubor `index.php` obsahuje jednotlivé funkce popsané ve vývojovém diagramu obrázek 3.2. Například funkce `function getIP` vrací IP adresu stanice komunikující s webovým serverem, funkce `function ip_is_private` vrací pravda, pokud je IP adresa z privátního rozsahu, funkce `function getMaxmind` vrací pole s názvem databáze, detekovanou zemí, městem s předpokládanou polohou a pokud jsou zadány správné souřadnice, tak i vypočítanou chybu lokace.

V prvním kroku je zjišťováno, zda formulář obsahuje nějaká data. Aby aplikace byla co nejvíce intuitivní a nebylo nutné zadávat IP adresu pro geolokaci, je automaticky detekována IP adresa přistupujícího klienta na server. Pokud formulář obsahuje název města nebo státu, jsou tyto údaje převedeny na souřadnice. Následně se zjišťuje, zda zadaná IP adresa nebo doména je platná. Pokud tento krok selže, je proces

zjišťování polohy zastaven a je vypsána hláška `Domain xxx does not exist`. Následující krok generuje dotazy na geolokační databáze. Na které geolokační servery zasílat požadavky, je možné vybrat zaškrtnutými políčky. Následně probíhá ověřování odpovědí. Odpovědi vyhodnocené jako neplatné jsou ve výsledcích zahrnuty, ovšem s výsledkem `Not available` a dále se tyto výsledky nepovažují za platné. Do aplikace lze zadat správnou lokaci aplikace dvěma způsoby. Za prvé jako ulici s číslem popisným a město. Tyto údaje jsou automaticky převedeny na souřadnice. Nebo lze zadat do aplikace přímo souřadnice. Jsou-li známy souřadnice, je vypočítána odchylka od lokalizované pozice v kilometrech. Jinak je k zadání souřadnic uživatel vyzván hláškou `Insert correct location`. V posledním kroku jsou data uspořádána do tabulky a je vygenerována mapa s popisem.

3.1 Popis programu

3.1.1 Vstupní parametry

Mezi povinný vstupní parametr patří IP adresa, bez které nelze realizovat měření. Aby chování aplikace bylo pro uživatele co nejintuitivnější, je IP adresa komunikujícího počítače detekována automaticky funkcí `getIP()`. Aby bylo možné provést geolokaci, musí být IP adresa z veřejného rozsahu. Kontrola je provedena pomocí funkce `ip_is_private()`. Je-li vyhodnoceno, že IP adresa z veřejného rozsahu je měření provedeno s touto IP adresou. Další parametry jsou nepovinné, jedná se o skutečnou adresu lokalizované adresy a souřadnice. Tyto údaje můžeme zadat dvěma způsoby a to jako adresu, kterou aplikace převede na souřadnice nebo jako samotné souřadnice. Z těchto údajů je vypočítána chyba lokace.

IP address or domain name to locate:			
Your IP address and domain name is:	78.102.208.117		
Correct address/cordinates to calculate location error			
Street:	<input type="text"/>	Latitude:	<input type="text"/>
City:	<input type="text"/>	Longitude:	<input type="text"/>
Select Geo Database: <input checked="" type="checkbox"/> SkyHook Wireless <input checked="" type="checkbox"/> DB-IP.com <input checked="" type="checkbox"/> Geobytes.com <input checked="" type="checkbox"/> Maxmind Lite <input checked="" type="checkbox"/> IP Info <input checked="" type="checkbox"/> FreeGeoIP <input checked="" type="checkbox"/> HostIP <input checked="" type="checkbox"/> IP2location <input checked="" type="checkbox"/> Whois			
Select additional information: <input checked="" type="checkbox"/> Delay <input checked="" type="checkbox"/> Proxy			
<input type="button" value="Find location"/>			

Obr. 3.1: Ukázka zadávání vstupních parametrů.

3.1.2 Převod ulice a města na souřadnice

Skutečnou fyzickou adresu aplikace převádí na souřadnice. K tomu je využito Google Decoding API, které vrací informace ve formátu JSON s předem známou strukturou dokumentu. Požadavek zadáváme ve formátu:

```
http://maps.google.com/maps/api/geocode/json?address=
```

Do řetězce je doplněna zjišťovaná adresa, která je ještě před odesláním zba-vena diakritiky funkcí `iso2ascii()`. Adresa může být zadána ve tvaru Technická 12, Brno-Královo Pole. Souřadnice se v odpovědi nacházejí v `{'results'}[0]->{'geometry'}->{'location'}->{'lat'}` a `{'results'}[0] ->{'geometry'} ->{'location'} ->{'lng'}`.

3.1.3 Výpočet odchyly

Po nalezení souřadnic z geolokačních databází a získání souřadnic vztažného bodu buď přímým zadáním nebo přepočtem z fyzické adresy, je dalším krokem výpočet geolokační chyby. K nalezení vzdálenosti mezi dvěma body tvořenými zeměpisnými souřadnicemi vede více možností. Mezi nejrozšířenější algoritmy patří Haversinův a Vincentův algoritmus. Země, z důvodu působení různých sil, nemá ideální tvar. Mnohdy se považuje za kouli stejně jako v případě Harvesinova algoritmu. Tvar Země se nejvíce blíží Geoidu, což je vlastně model povrchu Země při střední hladině oceánů. Považovat Zemi v rámci zjednodušení za kouli samozřejmě lze, ale dnes již výpočty nejsou prováděny ručně, ale hodnoty jsou počítány velmi výkonnými stoji. Proto program využívá složitějšího, ale přesnějšího Vincentova algoritmu.

Tento algoritmus je v geodezii využíván již od roku 1975. Předpokladem je, že Země má tvar rotačního elipsoidu, neboli elipsoidu, který má dvě poloosy shodné a je na pólech zploštělý. Není to zcela ideální model pro popis Země, ale je přesnější než koule. Jako nevýhodu lze uvažovat dvojnásobnou dobu výpočtu v porovnání s Harvesinovým algoritmem. Výhodou je přesnost, která je vyšší o několik řádů. Vzorec pro výpočet vzdálenosti mezi dvěma body Vincentovým algoritmem [16]:

$$\gamma = \arctan\left(\frac{a}{b}\right),$$

$$a = \sqrt{(\cos lat_1 \cdot \sin \Delta lat)^2 + (\cos lat_2 \cdot \sin lat_1 - \sin lat_2 \cdot \cos lat_1 \cdot \cos \Delta long)^2},$$

$$b = \sin(lat_2) \cdot \sin(lat_1) + \cos(lat_2) \cdot \cos(lat_1) \cdot \cos \Delta long, \quad (3.1)$$

$$d = R \cdot \gamma[\text{m}].$$

Kde:

R poloměr Země = 6371 km,

$\Delta long$ rozdíl mezi zeměpisnou šířkou bodu A a B.

Hodnota d je výslednou odchylkou udávanou v metrech mezi skutečnou polohou a lokalizovanou polohou. V aplikaci je výpočet chyby lokalizace uveden ve funkci `locationFault()`.

3.1.4 Získání dat z geolokačních databází

Na online databáze je zasláné specifické URI (Uniform Resource Identifier) obsahující IP adresu a u některých databází také uživatelské jméno, heslo nebo token. Jako odpověď je přijat JSON a nebo HTML dokument se všemi informacemi o dané IP adrese. Každá databáze vrací informace různým způsobem, bylo tedy nutné pro každou databázi použít rozdílný přístup a upravit data do jednotného výstupu, tak aby bylo možné data v rozumné formě prezentovat. U offline databáze je použita přímo naprogramovaná knihovna od Maxmind pro efektivní práci s databází.

IP Info

Tato online databáze vrací dokument JSON, ve kterém jsou obsaženy údaje o dotazované IP adrese. Souřadnice jsou obsaženy v jednom poli, odděleny pouze čárkou. K rozdělení souřadnic řetězce v poli a uložení do vlastních proměnných, je použita funkce v PHP `explode()`.

Následuje příklad výstupu databáze IP Info 3.1:

```
1 {
2   "ip": "77.75.72.3",
3   "hostname": "www.seznam.cz",
4   "city": null,
5   "region": null,
6   "country": "CZ",
7   "loc": "50.0833,14.4167",
8   "org": "AS43037 Seznam.cz, a.s."
9 }
```

Výpis kódu 3.1: Výstup databáze IP Info.

The DB-IP Database

Zde je situace podobná jako u online API (Application Programming Interface) Skyhook Wireless nejdříve je nutné získat API klíč, jinak API nevrací výsledky o dotazované IP adrese. Data jsou obsažena v JSON dokumentu, tudíž funkce k získání dat jsou stejné jako u předchozí databáze. Data jsou obsazena v poli ['country'] a ['city']. Komplikace nastává v případě, kdy chceme přesné souřadnice. Ty totiž ve volně dostupné verzi nejsou k dispozici, v placených verzích souřadnice standardně k dispozici jsou. Program pracuje pouze s názvem státu a názvem města. Tento nedostatek je řešen reverzním překladem názvu města na souřadnice přes Google API. Bohužel toto řešení se neukázalo jako funkční ve všech případech. A to proto, že občas nesouhlasí název města v databázi s databází Google API. Další komplikací je fakt, že Google API vrací polohu centra města, poloha tedy může být zkreslena.

Následuje příklad výstupu databáze DB-IP 3.2:

```
1 {  
2 "address": "212.79.109.151",  
3 "country": "CZ",  
4 "stateprov": "Vysočina Region",  
5 "city": "Nové Město na Moravě"  
6 }
```

Výpis kódu 3.2: Výstup databáze DB-IP.

Maxmind GeoLite2

S databází se pracuje offline přímo na hostitelském serveru. Jako zdroj dat byla vybrána databáze GeoLite2-City.mmdb vydávána každých 30 dní. Nahrávání nové verze lze provádět buď ručně nebo na Linuxových systémech pomocí cronu. K vyčítání dat je použita volně dostupná knihovna určena přímo pro PHP. Knihovna s názvem GeoIP2 PHP API funguje jak s Lite verzí databáze, tak i s placenou verzí. Samotné vyčítání dat z databáze se neobejde bez načtení knihovny příkazy `require_once 'vendor/autoload.php'` use `GeoIp2\Database\Reader` Dále se pracuje už se samotnou databází, která je v prvním kroku načtena příkazem `$reader = new Reader ('geoip2/GeoLite2-City.mmdb')`. Nyní je možné se dotázat na konkrétní IP adresu příkazem `$record = $reader->city($ip_addr)`. Je-li IP adresa v databázi nalezena, výsledek je vrácen jako pole a souřadnice jsou uloženy v: `$record->location->latitude` a `$record->location->longitude`.

Skyhook Wireless

Pro přístup do online API ve Skyhook Wireless je nejdříve nutné získat přístup do jejich databáze. Generované URI mimo dotazované IP adresy musí obsahovat uživatelské jméno a klíč. API poté vrátí JSON dokument. Uložení dokumentu do proměnné je provedeno pomocí funkce `file_get_contents()`, následně je převedeme zpět do JSON funkcí `json_decode()`. Nyní jsou data připravena ke zpracování. Informace jsou uloženy v poli `['data']['location']['latitude']` a `['data']['location']['longitude']`. Před samotným zpracováním je ověřeno, zda odpověď je pro stejnou IP adresu, na kterou bylo dotazováno.

Následuje příklad výstupu databáze Skyhook Wireless 3.3:

```
1 {
2   "data":{
3     "location":{
4       "type":"FIXED",
5       "latitude":49.209320068359375,
6       "longitude":16.603639602661133,
7       "hpe":2035.0
8     },
9     "ip":"78.102.208.117",
10    "civic":{
11      "state":"Jihomoravský kraj",
12      "country":"Czech Republic",
13      "city":"Brno",
14      "countryProb":1.0,
15      "countryIso":"CZ",
16      "stateProb":1.0,
17      "stateIso":"JM",
18      "cityProb":1.0
19    }
20  }
21 }
```

Výpis kódu 3.3: Výstup databáze Skyhook Wireless.

Geobytes

Do online databáze Geobytes bylo z počátku nutné vytvořit uživatelský účet a po jeho aktivaci byl umožněn omezený přístup k databázi. Databáze nabízela zdarma 20 dotazů za hodinu a ve chvíli, kdy počet dotazů byl překročen, přístup byl zablokován. Přístup bylo povoleno po nabití kreditu, ze kterého se odečítalo podle počtu dotazů. V průběhu semestru došlo ke změnám, databáze se stala bezplatnou s neomezeným počtem dotazů. Dotazy dříve byly vráceny ve formě meta tagů. Nyní jsou data dostupné v JSON. Ukládání do proměnné probíhalo funkcí `get_meta_tags()`. Nyní se pracuje s funkcemi `file_get_contents` a `json_decode`. Získaná data jsou uložena v polích. Například souřadnice jsou uloženy v poli `$response['latitude']` a `$response['longitude']`.

Příklad původního výstupu aplikace Geobytes 3.4:

```
1 <meta name="known" content="true">
2 <meta name="locationcode" content="CZJMBRNO">
3 <meta name="fips104" content="EZ">
4 <meta name="iso2" content="CZ">
5 <meta name="iso3" content="CZE">
6 <meta name="ison" content="203">
7 <meta name="internet" content="CZ">
8 <meta name="countryid" content="64">
9 <meta name="country" content="Czech Republic">
10 <meta name="regionid" content="5122">
11 <meta name="region" content="Jihomoravsky Kraj">
12 <meta name="regioncode" content="JM">
13 <meta name="adm1code" content="EZ78">
14 <meta name="cityid" content="3639">
15 <meta name="city" content="Brno">
16 <meta name="latitude" content="49.2000">
17 <meta name="longitude" content="16.6330">
18 <meta name="timezone" content="+01:00">
19 <meta name="certainty" content="77">
20 <meta name="mapbytesremaining" content="Free">
```

Výpis kódu 3.4: První verze výstupu databáze Geobytes.

Následuje aktuální výstup databáze Geobytes, který využívá jako ostatní výstupy, formát JSON 3.5:

```

1 {
2 "geobytesinternet":"US",
3 "geobytescountry":"United States",
4 "geobytesregionlocationcode":"USAZ",
5 "geobytesregion":"Arizona",
6 "geobytescode":"AZ",
7 "geobyteslocationcode":"USAZTUCS",
8 "geobytescity":"Tucson",
9 "geobytescityid":"7056",
10 "geobytesfqcn":"Tucson, AZ, United States",
11 "geobyteslatitude":"32.232700",
12 "geobyteslongitude":"-110.945999",
13 "geobytescapital":"Washington, DC ",
14 "geobytestimezone":"-110.946",
15 "geobytesnationalitysingular":"American",
16 "geobytespopulation":"278058881",
17 "geobytesnationalityplural":"Americans",
18 "geobytesmapreference":"North America ",
19 "geobytescurrency":"US Dollar",
20 "geobytescurrencycode":"USD",
21 "geobytestitle":"The United States"
22 }

```

Výpis kódu 3.5: Aktuální výstup databáze Geobytes.

FreeGeoIP, HostIP a IP2Location

Jedná se o volně dostupné databáze, které spravuje kolega Janoušek. Data si mezi weby předáváme pomocí JSON. Aby nedocházelo ke zbytečnému dotazování stále dokola, jsou data o databázích přenášena společně v jednom JSON dotazu a dále jsou rozdělena podle jednotlivých databází a zobrazena tak, jako by se jednalo o dotazy na jednotlivé databáze. JSON data lze získat po zadání následující adresy:

<http://geolocation.utko.feec.vutbr.cz/xjanou09/?ip=77.75.72.3>

Výstup byl navržen tak, aby bylo možné co nejjednodušeji vyčítat data. Veškerá data jsou obsažena v poli data a dále jsou dělena ve skupinách. Skupina 0 obsahuje data z databáze FreeGeoIP. Skupina 1 obsahuje získaná data z HostIP a poslední skupina číslo 3 obsahuje získaná data z DB-IP. Mimo tyto skupiny dat je ještě v JSON obsažená IP adresa pro kontrolu, zda došlo ke správnému dotazu. Každá

skupina dat, tedy jedna geolokační databáze, obsahuje položky database, kde je uveden název databáze, country, kde je uveden kód země, region obsahuje informace o kraji, zeměpisná šířka a výška je uvedena v poli latitude a longitude, znovu je zde uvedena dotazovaná IP adresa v poli ipaddress a jako doplňující je zde pole error, které se momentálně nepoužívá. Příklad výstupu 3.6:

```
1 {
2   "data": [
3     {
4       "database": "Freegeoip",
5       "country": "CZ",
6       "region": "",
7       "city": "",
8       "latitude": 50.083,
9       "longitude": 14.417,
10      "ipaddress": "77.75.72.3",
11      "error": false
12    },
13    {
14      "database": "HostIP",
15      "country": "CZ",
16      "region": "",
17      "city": "Prague",
18      "latitude": "50.0833",
19      "longitude": "14.4333",
20      "ipaddress": "77.75.72.3",
21      "error": false
22    },
23  ],
24  "ip": "77.75.72.3",
25  "lat": "",
26  "lng": ""
27  },
28  {
29    "database": "DB-IP",
30    "country": "CZ",
31    "region": "Prague",
32    "city": "Prague",
33    "latitude": 50.0755381,
34    "longitude": 14.4378005,
35    "ipaddress": "77.75.72.3",
36    "error": false
37  },
38  {
39    "database": "IP2location",
40    "country": "CZ",
41    "region": "HLAVNI MESTO PRAHA",
42    "city": "PRAGUE",
43    "latitude": 50.088039398193,
44    "longitude": 14.420760154724,
45    "ipaddress": "77.75.72.3",
46    "error": false
47  },
48  ],
49  "ip": "77.75.72.3",
50  "lat": "",
51  "lng": ""
52  }
```

Výpis kódu 3.6: Výstup dotazu na databáze FreeGeoIP, HostIP a IP2Location.

Měření zpoždění

Zpoždění není měřeno přímo na webové stránce. Jedná se o data, která jsou měřena aplikací kolegy pana Bednáře. Měření je vyhodnoceno jako neplatné pokud odpověď na ICMP paket (Echo Request) není přijata za déle než 1 000 ms. Čas je zvolen tak, aby v případě, kdy IP adresa neodpovídá, bylo načítání stránky stále rychlé. Je-li první měření úspěšné, ping je proveden ještě jedenkrát. Výsledek je před vložením do JSON zprůměrován. Doba čekání na odpověď ani počet pokusů není zcela ideální. Dle mého názoru, aby se předešlo nepřesným výsledkům, by bylo vhodné měření provést minimálně 5x za sebou s pauzou alespoň 500 ms. Měření by pak trvalo přibližně 2 500 ms. Dále, když započteme generování JSON předávání dat a vygenerování stránky, dostáváme se k 3 000 ms, které jsou již kriticky dlouhé. Dotaz na měření zpoždění je odeslán ve tvaru:

```
http://geolocation.utko.feec.vutbr.cz/xbedna54/?ip=77.75.72.3
```

Data jsou vracena ve formátu JSON a vrací se dotazovaná IP adresa pouze pro kontrolu, zda nedochází při zpracování k chybě. Položka `delay` nabývá dvou typů. Textová hodnota `not available` je použita tehdy, kdy je měření neúspěšné. Druhým typem je číselná hodnota naměřeného zpoždění. Formát JSON odpovědi 3.7:

```
1 {  
2 "address": "77.75.72.3",  
3 "delay": "5.291"  
4 }
```

Výpis kódu 3.7: Výstup dotazu na měření zpoždění.

Detekce pomocí WHOIS databáze

Vyhledání lokace pomocí WHOIS je další externí funkcí, kterou využívá vytvořená aplikace. Data jsou získána od pana Smrčky ve formátu JSON. Nejdříve zde bude popsáno, jak lze obecně získat údaje z databáze WHOIS. Po zadání hledané IP adresy je nejprve zjištěno hostname, které je v následujícím nebo podobném tvaru např.: `t151.steadynet.cz`. Tento konkrétní hostname můžeme rozdělit na domény prvního až třetího řádu. Pro vyhledávání v WHOIS záznamech je nejdůležitější doména druhého řádu `steadynet.cz`, která ve zmiňovaném případě určuje poskytovatele internetového připojení.

WHOIS obsahuje různá data o registrátorovi. Z hlediska geolokace nás zajímají pole `address`. Pro zmiňovanou IP adresu obsahuje následující data na 3.8:

```
1 address: Vratislavovo namesti 190
2 address: Nove Mesto na Morave
3 address: 59231
4 address: CZ
```

Výpis kódu 3.8: Zkrácený výstup databáze WHOIS.

Problematika získávání dat z WHOIS databáze je komplikovanější, než je uvedeno na příkladu, zejména v případě, kdy nejsou obsažena pole `address`. Údaje o adrese registrátora mohou být rozptýleny do jiných částí WHOIS záznamu.

Využívané API od pana Smrčky, které získává informace z WHOIS databází, filtruje užitečná data a převádí získanou adresu na souřadnice, lze nalézt po zadání adresy ve tvaru:

```
http://geolocation.utko.feec.vutbr.cz/
xsmrck00/search.php?json=true&ip=212.79.109.151
```

Data jsou ve formátu JSON vrací se pole `ip` pro kontrolu, zda opravdu došlo k detekci správné IP adresy. Následující pole `country` obsahuje kód státu. Další pole `city` obsahuje název města. Pole `street` obsahuje název ulice. Následuje pole souřadnic `latitude` a `longtitude`. Jak se převádí fyzická adresa na souřadnice, je uvedeno v kapitole 3.1.2. Pole `NOTE` je využíváno pro poznámku, která bude využívána v dalších verzích chystaných v následujících letech. Vracený JSON 3.9 je pro výše zmíněnou IP adresu následující:

```
1 {
2   "ip": "212.79.109.151",
3   "country": "CZ",
4   "city": "Nové Město na Moravě",
5   "street": "Not located",
6   "latitude": "49.5614432",
7   "longtitude": "16.0741877",
8   "NOTE": "592 31 Nove Mesto na Morave"
9 }
```

Výpis kódu 3.9: Výstup dotazu na filtrovaná data z databáze WHOIS.

Detekce proxy

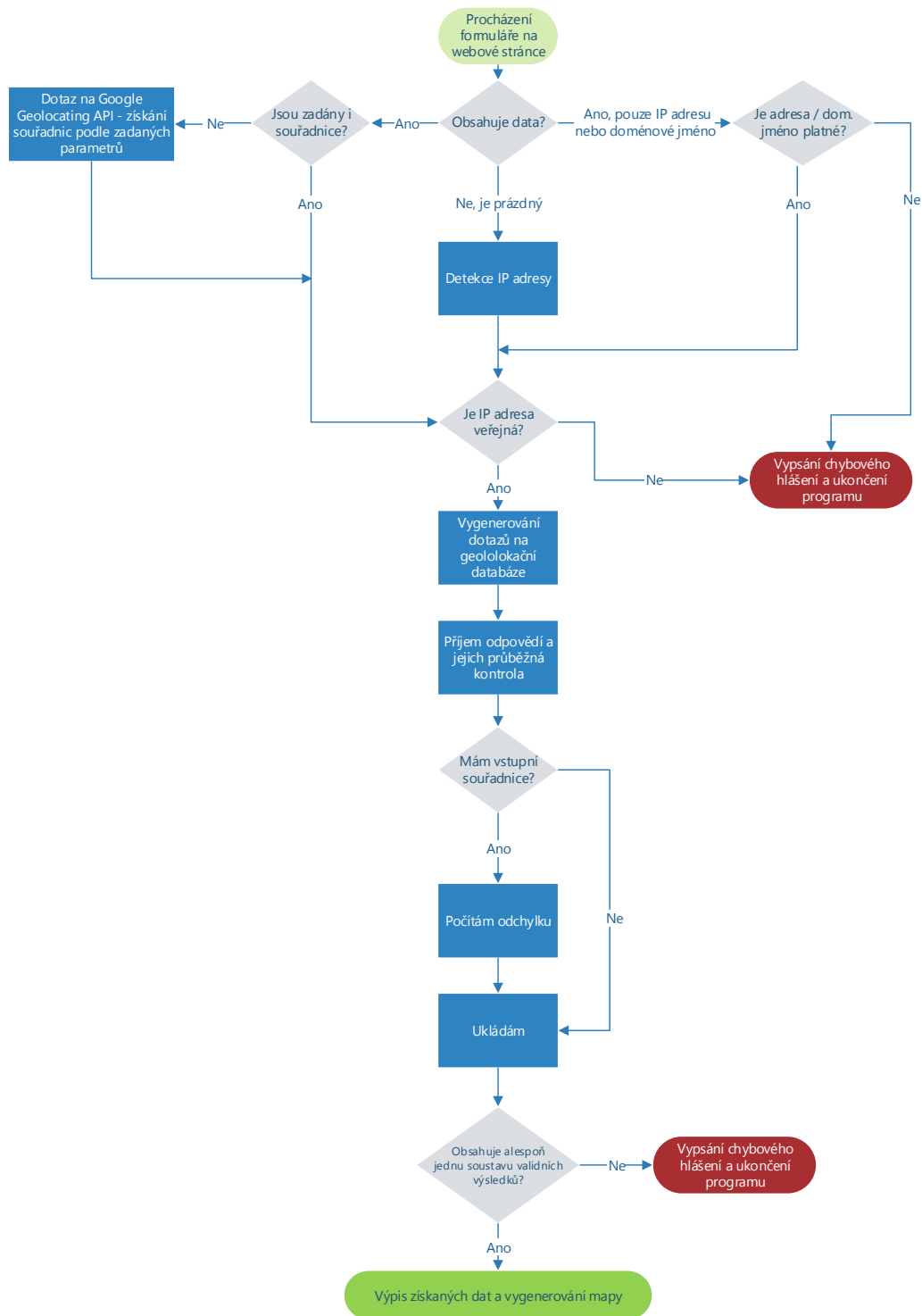
Data o přítomnosti proxy serveru jsou získávána z aplikace pana Gardiana, který má bakalářskou práci na toto téma a poskytl výstup ze své aplikace ve formátu JSON pro využití na tomto webu. Ve velmi jednoduchém principu jde o prohledání veřejně dostupných seznamů IP adres, které jsou známé jako proxy servery. Pokud je IP adresa v seznamu nalezena, je označena jako proxy server. V prvních fázích vývoje docházelo k prohledávání seznamů IP adres při každém dotazu. Takový postup se ukázal jako velmi časově neefektivní a mnohdy došlo k přerušení komunikace ze serverem. Standardně je doba vykonávání příkazu v PHP 30 sekund poté je komunikace přerušena. Z toho důvodu je možné, že k nalezení odpovědi nemusí vždy dojít. Řešením je na serveru nastavit vyšší timeout.

Odpověď s daty je získávána ve formě JSON a lze získat po zadání následující adresy do webového prohlížeče:

```
http://geolocation.utko.feec.vutbr.cz/xgardi01/  
api.php?ip=212.79.109.151
```

Odpověď obsahuje pouze IP adresu pro kontrolu, zda byl dotaz vygenerován správně a jako další položku obsahuje vlastní pole proxy, které nabývá dvou hodnot a to **true** nebo **false**.

Odpověď je zpracována dále do tabulky a je zobrazena pod mapu detekovaných míst pomocí geolokačních databází. Společně s detekcí proxy je ve stejné tabulce zobrazeno také naměřené zpoždění.



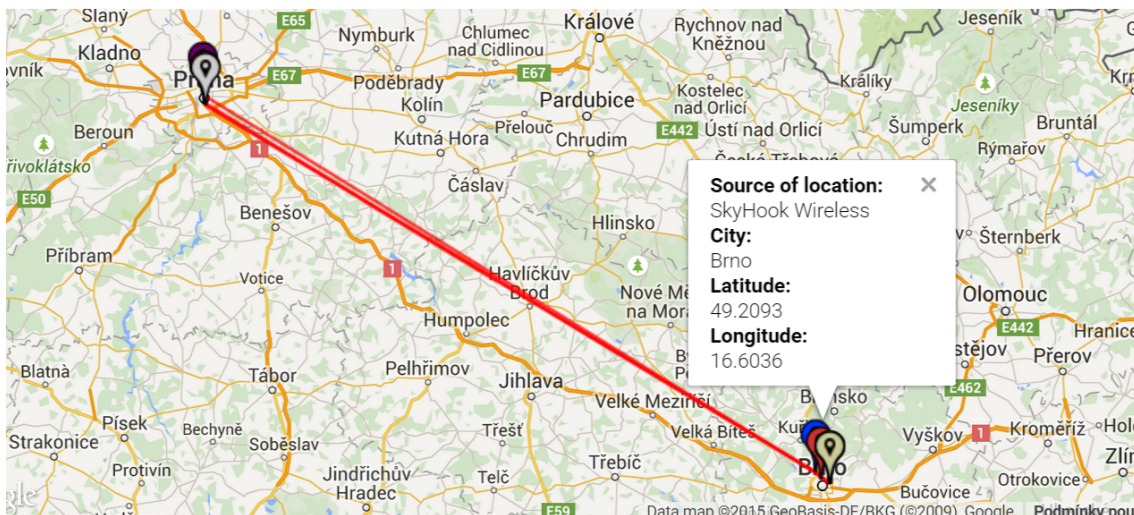
Obr. 3.2: Vývojový diagram vytvořené aplikace.

3.1.5 Grafická prezentace výsledků

Výsledky jsou přehledně prezentovány do tabulky tak, aby bylo možné přidávat nebo ubírat počet využívaných databází bez nutnosti velkých úprav kódu. Generování tabulky zajišťuje funkce `generateTable`. V prvním sloupci je název geolokační databáze, následuje detekovaná země. Ve třetím sloupci je detekované město a následují souřadnice. V poslední sloupci je vypočítána chyba geolokace.

Location database	Country	City	Latitude	Longitude	Location error [km]
SkyHook Wireless	CZ	Brno	49.2093	16.6036	2.29
DB-IP.com	CZ	Prague	50.0624463	14.4404548	180.19
Geobytes.com	CZ	Brno	49.2	16.633	4.57
Maxmind Lite	CZ	Brno	49.2	16.6333	4.59
IP Info	CZ	Brno	49.2	16.6333	4.59
Whois	CZ	Praha	50.0582	14.4443	179.72
FreeGeolP	CZ	Brno	49.2	16.633	4.57
HostIP.info	CZ	Prague	50.0833	14.4333	181.81
IP2Location	CZ	Prague (Prague 4)	50.0624	14.4405	180.19

Obr. 3.3: Ukázka výsledků v tabulce.



Obr. 3.4: Ukázka výsledků v mapě.

Získané souřadnice jsou graficky zobrazeny v mapě. Aplikace využívá Google Maps API, které nabízí vkládání vlastních bodů tzv. „markers“. Každý takový bod lze individuálně modifikovat a po kliknutí na tento bod lze zobrazit doplňující informace. Data se vkládají do API pomocí Javascriptu. Google Maps API je nejdříve

staženo v hlavičce stránky a následně je upravováno doplňujícím kódem po dokončení geolokace. Do mapy je vyznačena jen databáze, která má informace o geografické poloze stanice. Každý takový bod je odlišen barevně. Po kliknutí na bod jsou zobrazeny doplňující informace, jedná se o název geolokační databáze, ze které jsou data získány, město, zeměpisná šířka a zeměpisná délka. Google API nenabízí automatické centrování a automatický zoom na vypsání body. Před samotným zobrazením mapy je vypočítán průměr ze získaných souřadnic a nastaveno zvětšení mapy. Pokud je vzdálenost mezi body větší než 500 km, je nastaven maximální zoom, aby byly všechny body viditelné.

3.1.6 Ukládání a zobrazení naměřených výsledků

Každé měření se zadanou polohou je ukládáno do textového souboru umístěného v `/data/locations.txt`. První podmínkou pro uložení měření je platná IP adresa z veřejného rozsahu získána detekcí klientovi IP adresy, vložením IP adresy nebo úspěšným překladem z doménového jména. Druhou podmínkou je zadání souřadnic nebo úspěšný překlad z názvu ulice a města na souřadnice. Jeden řádek v textovém souboru reprezentuje jedno měření, které obsahuje lokalizovanou IP adresu, vložené souřadnice a poté následují jednotlivé databáze se svým jménem a detekovanými souřadnicemi. Na úvodní webové stránce je odkaz, který zobrazuje veškerá uložená data přehledně do tabulek obr. 3.5. Výsledky jsou uloženy přesně tak, jak byly zobrazeny uživateli. Pokud databáze nebyla schopna IP adresu lokalizovat, je zobrazeno `Not available`. Nad tabulkou se nachází odkaz, který zobrazuje data z vybrané tabulky do mapy a zároveň je zobrazuje pro přehlednost znovu do tabulky. Zadaná pozice je spojena s lokalizovanými body pro jednodušší orientaci přímkou, obrázek 3.6. Jedno měření je uloženo jako jeden řádek, zde je vysázeno se zalomením pro přehlednost 3.10. Výpis začíná lokalizovanou IP adresou, souřadnicemi, dále pak název databáze s detekovanými souřadnicemi.

```
1 147.229.93.23,49.2260359,16.5755884
2 SkyHook Wireless,49.2306,16.67,6.87
3 DB-IP.com,49.1924655,16.605168,4.31
4 Geobytes.com,49.683,18.333,136.81
5 Maxmind Lite,49.2,16.6333,5.09
6 IP Info,49.2,16.6333,5.09
7 FreeGeoIP,49.2,16.633,5.08
8 HostIP.info,49.2,16.6333,5.09
9 IP2Location,49.1925,16.6052,4.31
```

Výpis kódu 3.10: Uložené informace o IP adrese a její poloze v souboru `location.txt`.

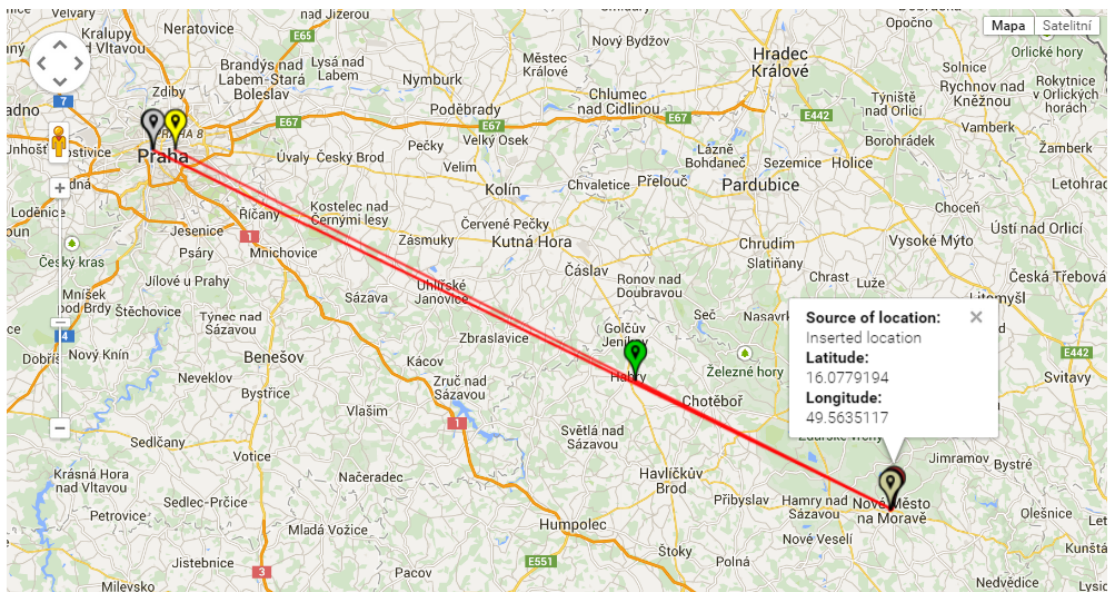
IP address: 212.79.109.151
 Inserted location: 49.5635117 16.0779194
[Click to show results on map](#)

Location database	Latitude	Longitude	Location Error [km]
SkyHook Wireless	49.5664	16.0817	0.42
DB-IP.com	49.5614432	16.0741877	0.35
Geobytes.com	50.083	14.467	129.19
Maxmind Lite	49.75	15.5	46.48
IP Info	50.0833	14.4167	132.44
FreeGeolP	50.083	14.417	132.41
HostIP.info	Not avaiable	Not avaiable	Can not be calculted
IP2Location	49.5614	16.0742	0.35

Obr. 3.5: Výpis naměřených výsledků.

IP address: 212.79.109.151
 Entered Location: 49.5635117 16.0779194

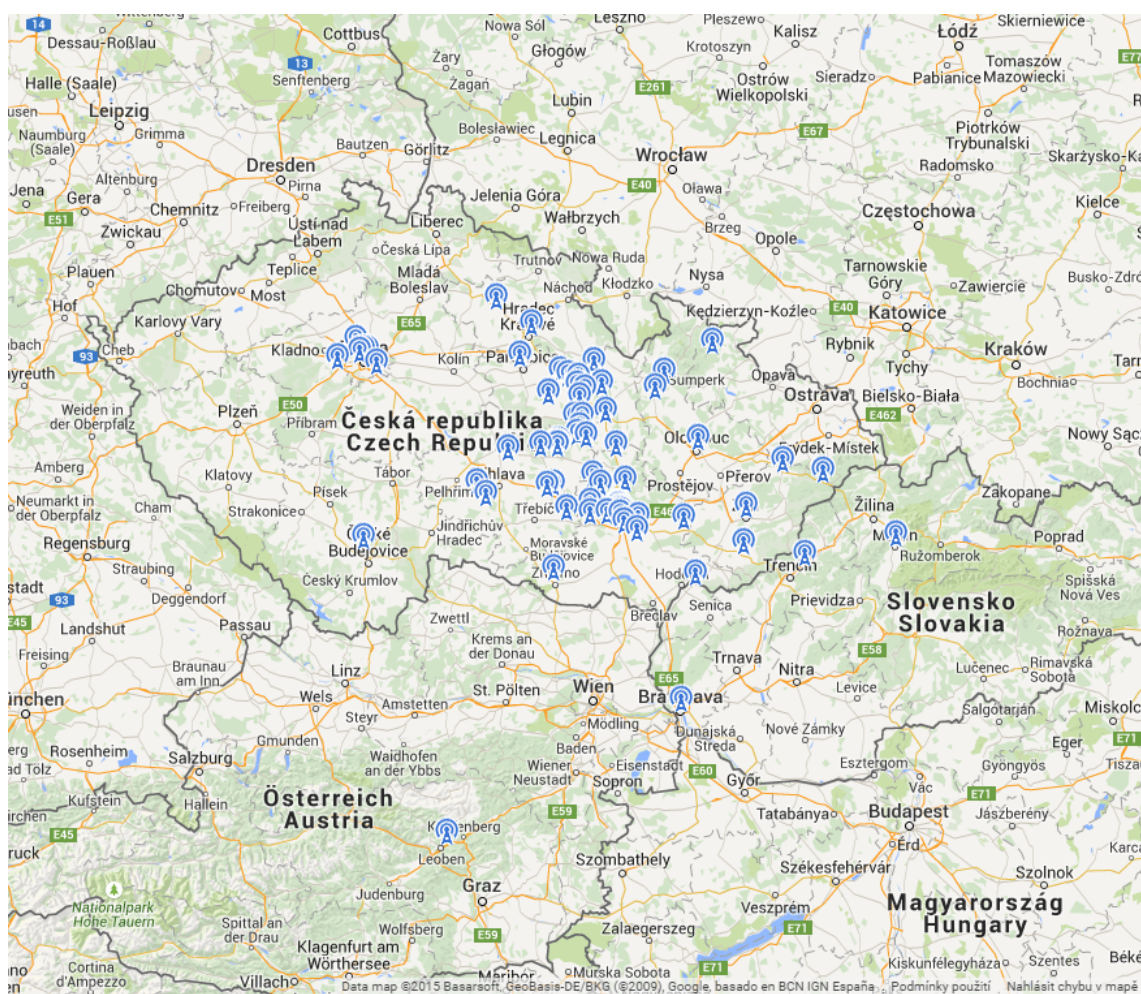
Location database	Latitude	Longitude	Location Error [km]
SkyHook Wireless	49.5664	16.0817	0.42
DB-IP.com	49.5614432	16.0741877	0.35
Geobytes.com	50.083	14.467	129.19
Maxmind Lite	49.75	15.5	46.48
IP Info	50.0833	14.4167	132.44
FreeGeolP	50.083	14.417	132.41
HostIP.info	Not avaiable	Not avaiable	Can not be calculted
IP2Location	49.5614	16.0742	0.35



Obr. 3.6: Zobrazení historie výsledků do mapy.

4 POROVNÁNÍ VÝSLEDKŮ PLACENÝCH GEOLOKAČNÍCH DATABÁZÍ

Aby bylo možné porovnat přesnost geolokačních databází s co největší přesností, bylo nutné získat co největší databázi IP adres se známou polohou. Podařilo se získat 81 IP adres se známou polohou, což považuji za dostatečný vzorek. Všechny IP adresy pocházejí z evropského kontinentu. Bohužel se nepodařil, získat dostatečný vzorek IP adres i z jiných světadílů. Velká většina IP adres pochází buď z České republiky nebo Slovenské republiky. Jejich konkrétní umístění je zobrazeno na obrázku 4.1. Další IP adresy se také nacházely v Irsku a v Turecku, pro lepší názornost nejsou v mapě zahrnuty. Souřadnice IP adres byly zjištěny zaměřením polohy přes GPS přijímač nebo přepočtem známé polohy na souřadnice přes mapový portál Google. Aktuální mapu s polohou všech zadaných IP adres do aplikace naleznete na stránce: <http://geolocation.utko.feec.vutbr.cz/xmodra00/maponly.php>.



Obr. 4.1: Poloha lokalizovaných IP adres na mapě.

4.1 Zpracování výsledků geolokačních databází

Postupně bylo zadáno 81 IP adres a jejich poloha ve formě souřadnic do naprogramované webové aplikace. Souřadnice jsou považovány za skutečnou polohu a nezáleží, zda byly získány přímo zaměřením GPS souřadnic nebo byly získány z mapy. Geolokační databáze určily polohu a aplikace spočítala odchylku měření, následně byly data uložena do textového souboru pro zpracování. Soubor byl naimportován do programu Microsoft Excel, ve kterém jsou vytvořeny následující tabulky a grafy.

Naměřené hodnoty byly rozděleny podle jednotlivých geolokačních databází a poté na výsledky do skupin podle naměřené vzdálenosti od lokalizovaných bodů. První skupinou jsou výsledky do 10 km od lokalizovaného bodu. Dále jsou výsledky rozděleny do 50 km, 100 km, 250 km a 500 km. Nad 500km je výsledek považován za neplatný a je označen jako chyba lokace. Nebyla-li adresa nalezena v databázi, pak je označena jako nelokalizováno. Vztáhnout chybu lokace na jednotku vzdálenosti je vhodné zejména, když si chceme naměřené výsledky představit na mapě. Daleko zajímavější úhel pohledu nabízí srovnání správnosti detekce města. A to zejména u inzerentů, výše zmíněných internetových vyhledávačů, vyhledávání počasí na základě detekované polohy nebo dopravních spojů, je schopnost detekovat správně město velmi důležitou vlastností. Jelikož detekce přesnosti na město nedosahuje přílišné přesnosti, byla přidána také přesnost detekce kraje, která se ukazuje například v zobrazování předpovědi počasí jako dostatečná.

4.1.1 Průměrná odchylka databází

Tabulka 4.1 ukazuje vypočtenou průměrnou odchylku od správné polohy IP adresy, kde není započtena chyba větší než 500 km. Tím získává výhodu geolokační server Geobytes.com, který jako jediný naměřil více než jednu IP adresu s odchylkou větší než 500 km a to více než v 21 %. I přes korekci naměřil v průměru nejvyšší vzdálenost od správné polohy ze všech databází. Nejmenší průměrnou odchylku naměřila databáze SkyHook Wireless s průměrnou odchylkou 36 km, která patří k velmi dobrým výsledkům. Průměrná odchylka zbývajících čtyřech databází je 80 km.

Databáze	Naměřená odchylka [km]
SkyHook Wireless	36
DB-IP.com	78
Geobytes.com	94
Maxmind Lite	71
IP Info	76

Tab. 4.1: Průměrná naměřená odchylka.

Výsledek je natolik odlišný, že bylo vhodné najít příčinu a tou je vysoký počet nelokalizovaných adres. Některé nelokalizované adresy u databáze SkyHook Wireless se ukázaly být jako problematické i pro ostatní geolokační služby. Naměřené hodnoty pro nelokalizované IP adresy jsou uvedeny v tabulce 4.2.

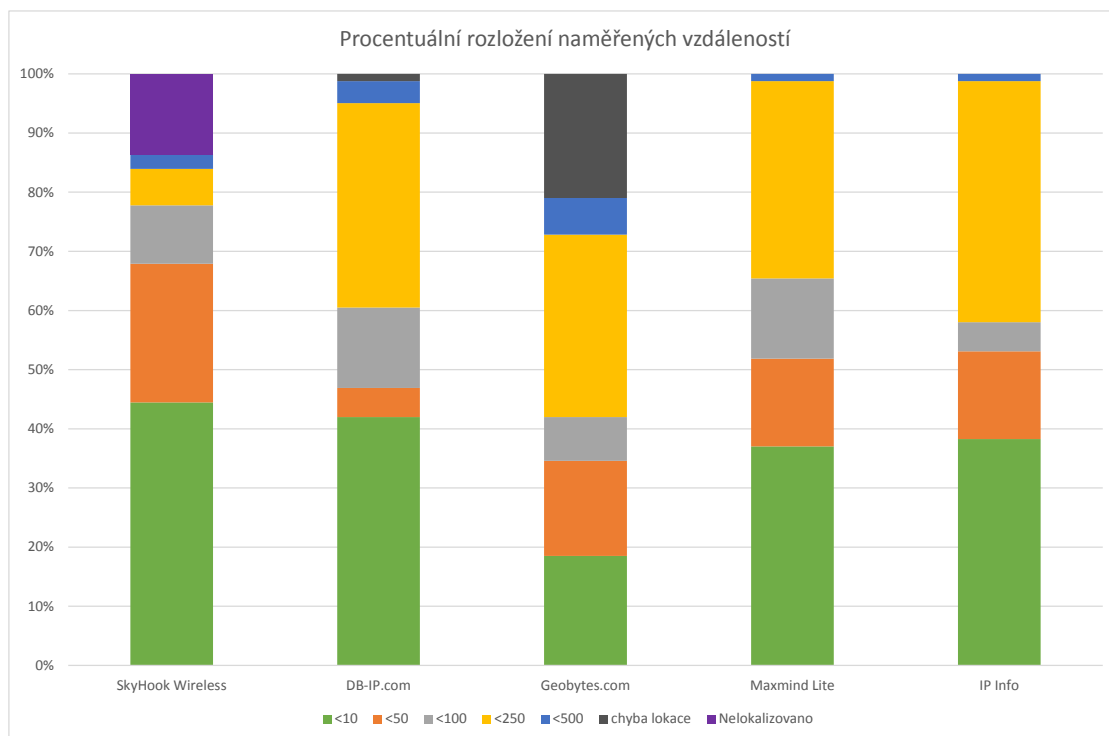
Z výsledků je patrné, že největší problém s nelokalizovanými IP adresami má databáze Geobytes.com, kdy šest z jedenácti adres můžeme označit za nesprávně zaměřené. Naopak nejlepších výsledků hluboko pod svým průměrem dosáhla databáze Maxmind Lite, která naopak měla menší komplikace pouze s jedním výsledkem. Za to z osmi výsledky z jedenácti si poradila velmi dobře. S lepšími výsledky než je průměr si s měřením poradila databáze DB-IP.com, pokud by byl zanedbán nejhorší naměřený výsledek. Databáze by dosáhla shodného výsledku jako Maxmind Lite. Databáze IP Info dosáhla shodného průměru při nelokalizovaných adresách jako je průměr databáze celého měření.

DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
0,27	8477,03	109,44	114,50
7,88	1012,20	54,67	132,36
4,01	46,42	51,18	134,69
262,50	34,93	0,04	0,04
0,53	93,04	2,31	2,31
5,81	1010,70	54,79	131,74
0,27	8477,03	109,44	114,50
209,07	152,2	211,76	211,76
0,7	7101,08	0,67	0,67
0,26	1,17	0,67	0,67
6565,87	8727,48	4,03	4,03

Tab. 4.2: Odchylka pro nelokalizované IP adresy databází SkyHook Wireless.

4.1.2 Procentuální rozložení naměřených vzdáleností

V grafu na obrázku 4.2 je zobrazeno procentuální rozložení vzdáleností v již zmíněných skupinách vzdáleností, kde 100 % jsou provedena všechna měření. Je zde velmi dobře vidět, že nejlépe lokalizuje IP adresy databáze Skyhook Wireless a to do vzdálenosti 10 km s úspěšností 44 %. Naopak nejhůře si v oblasti měření vede databáze Geobytes.com, která s chybou do 10 km lokalizovala celých 19 % IP adres. Ostatní databáze se pohybovali mezi 37 % až 42 %, což lze považovat za vyrovnaný výsledek.



Obr. 4.2: Statistika odchylek lokalizovaných IP adres.

Výsledky lokalizace mezi 10 km až 50 km lze bez delšího zkoumání považovat za přesné v rámci kraje (hlubší porovnání detekce města a krajů je uvedeno dále). Nejlepší výsledek 23 % v rozmezí 10 km až 50 km dosáhla databáze SkyHook Wireless. Naopak nejhoršího výsledku pouze 5 % dosáhla databáze DB-IP.com, která dosud dosahovala velmi slibných výsledků. Geobytes.com lokalizovalo 16 %, Maxmind Lite 15 % a IP Info také shodně lokalizovala 15 %.

Hledáme-li velmi přesné výsledky do 50 km, vychází nejlépe SkyHook Wireless s IP Info. Analýza nelokalizovaných IP adres od Skyhook Wireless ale ukázala jinou a mnohem přesnější variantu.

Databáze	Chyba do 50 km [%]
SkyHook Wireless	68
DB-IP.com	47
Geobytes.com	35
Maxmind Lite	52
IP Info	53

Tab. 4.3: Lokalizované IP adresy s chybou do 50 km.

Jelikož SkyHook do 50 km lokalizuje 68 % v případě, kdy nedojde k lokalizaci ze strany Skyhook Wireless, je výhodnější použít jako zdroj dat Mamind Lite, který

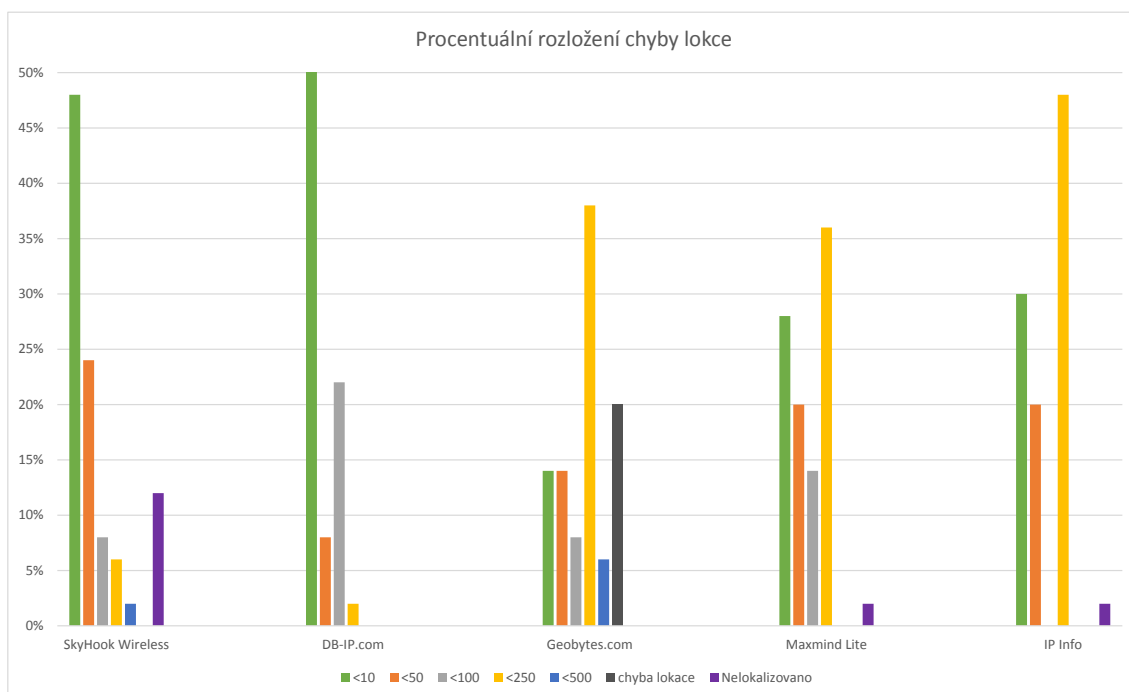
až na čtyři měření určit polohu s přesností do 55 km. Jako další varianta by mohla být kombinace databází Skyhook Wireless a DB-IP.com, která nepřesně lokalizovala pouze tři IP adresy, za to s větší chybou. Osm IP adres bylo lokalizováno s přesností do 10 km. IP Info lokalizovalo do 50 km 53 % IP adres a Maxmind Lite 52 %. U Geobytes došlo k mírnému zlepšení na 35 %, i přesto je databáze Geobytes nepoužitelná pro potřeby geolokace.

Nejvyšší procento lokalizovaných IP adres se paradoxně nachází v rozsahu mezi 100 až 250 km, takové rozlišení lze považovat přesné z pohledu určení státu. Z toho vyplývá, že geolokační databáze s výjimkou Skyhook Wireless nejsou použitelné pro detekci lokace IP adresy na úrovni města, i když takovou funkci všechny databáze nabízejí.

Skyhook Wireless	DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
6 %	35 %	33 %	33 %	41 %

Tab. 4.4: Odchylna v procentech pro chybu mezi 100 až 250 km.

Pro lepší představu o naměřených datech je na obrázku 4.3 zobrazena chyba lokace po jednotlivých kategoriích do 10 km, 10 km – 50 km, 50 km – 100 km, 100 km – 250 km, 250 km – 500 km a nad 500 km.



Obr. 4.3: Chyba lokace v zobrazení po jednotlivých kategoriích.

4.1.3 Vyhodnocení efektivity lokace

Přestože počet lokalizovaných IP adres není z pohledu adresního rozsahu mnoho, i tak se některé z nich nepodařilo lokalizovat nebo lokalizace nebyla provedena s očekávanou přesností. Výsledky v grafech na obrázku 4.4 jsou rozděleny na tři části. Adresy označené jako lokalizováno, jsou určeny s přesností do 500 km. Adresy označené jako chyba lokace, jsou adresy lokalizované, ale s chybou větší než 500 km a poslední částí je nelokalizováno. Tento stav nastává v případě, kdy databáze nemá údaj o lokalizované adrese ve svých tabulkách.

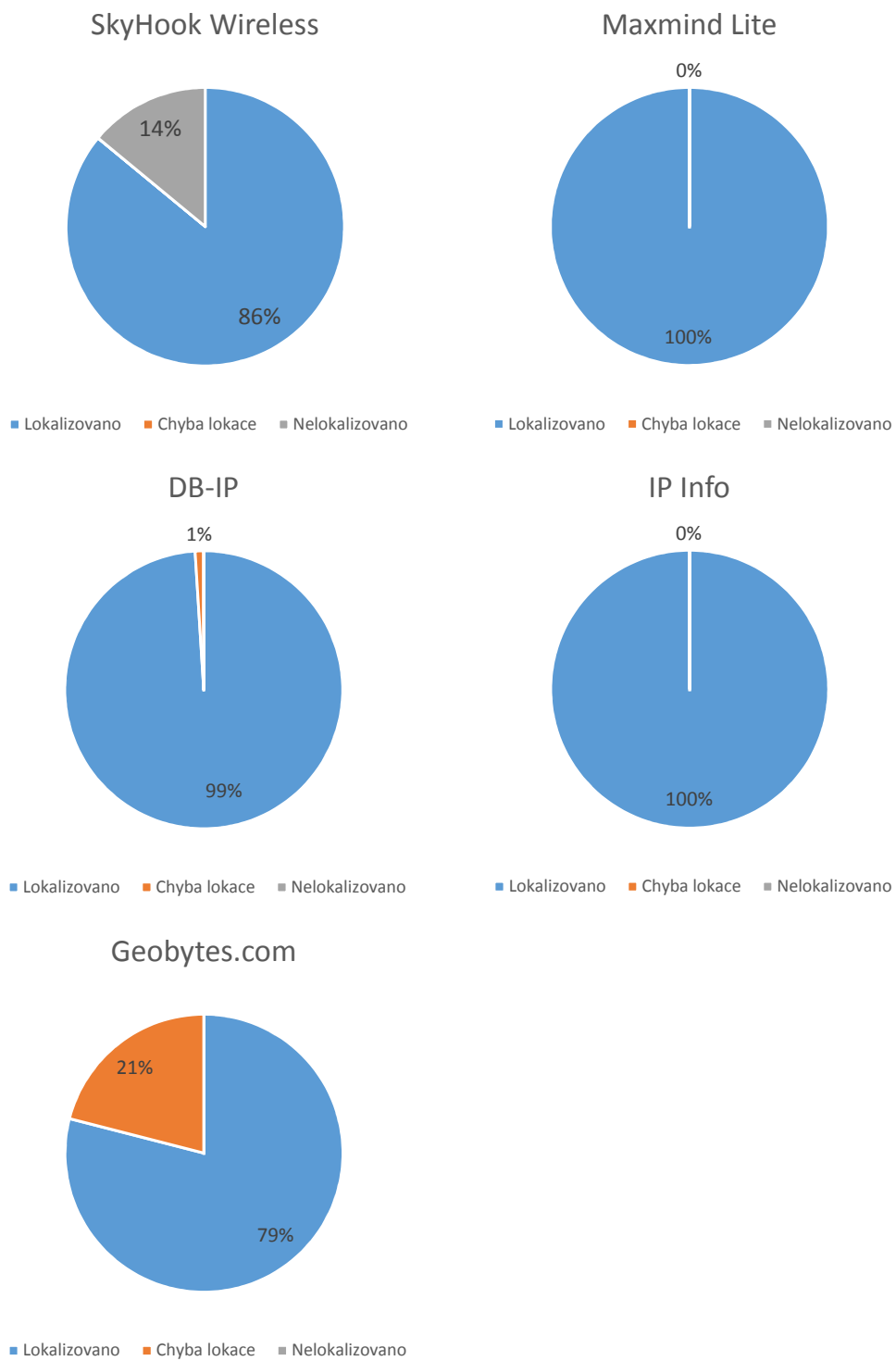
Z pohledu lokalizovaných IP adres si nejlépe vedli databáze Maxmind Lite a IP-Info, které dokázali přesně lokalizovat všechny zadané IP adresy. DB-IP nedokázala určit přesně jednu IP adresu a tak má úspěšnost lokalizace 99 %. Databáze Skyhook Wireless správně lokalizovala 86 % zadaných IP adres. Zbýlých 14 % IP adres nebylo v databázi nalezeno. Jak již bylo napsáno výše, jednalo se především o IP adresy, které měli jiné lokační databáze problém lokalizovat. Na tento výsledek se lze dívat tak, že pokud ve Skyhook Wireless neměli dostatek informací o poloze IP adresy, pak nebyla IP adresa do databáze zařazena vůbec. Nejhorší výsledek pouze 79 % lokalizovaných IP adres má databáze Geobytes.com, která napříč všemi vyhodnoceními má velmi špatné výsledky. Zbýlých 21 % IP adres bylo lokalizováno s chybou větší než 500 km. Nebylo výjimkou, kdy chyba lokace byla větší než 5 000 km. Jednalo se o IP adresy přidělené v České republice a jejich určená lokace databází Geobytes.com byla nejčastěji na území USA v Severní Americe.

Distribuční funkce

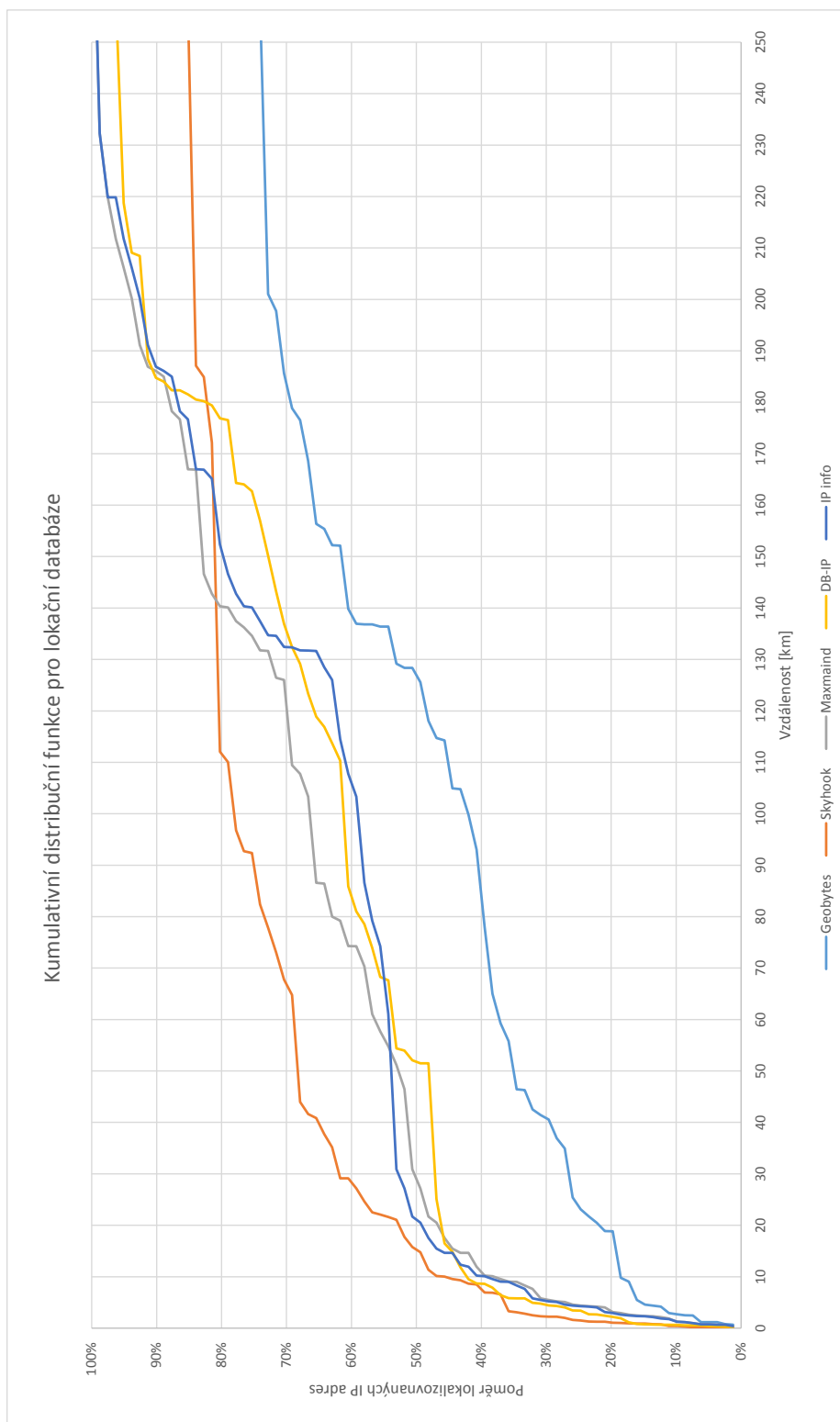
Pro úplnost na obrázku 4.5 je uvedena distribuční funkce. Jedná se o rozdělení pravděpodobnosti s jakou lze dosáhnout výsledku geolokace s požadovanou přesností. Jak je vidět z grafu, přesnost lokace do 17 km velmi prudce roste a pohybuje se mezi 42 % – 52 %. Výjimku tvoří databáze Geobytes, která dosahuje přesnosti kolem 18 %. S rostoucí chybou lokace se jednotlivé databáze srovnávají a s chybou lokace 185 km dojde k nejnižšímu rozptylu ve výsledcích přesnosti geolokace, kde všechny databáze se pohybují kolem 85 %. Mírně zde začíná zaostávat databáze Skyhook Wireless, která již dosáhla svého vrcholu a dále téměř neroste. Výjimkou je opět databáze Geobytes, která ve výsledcích zaostává o 15 %.

4.1.4 Přesnost lokačních databází na základě správních celků

Databáze nevrací pouze souřadnice, ale data obsahují hodnoty jako stát a město. Některé databáze obsahují informace o kraji, ve kterém se lokalizované místo nachází. Jelikož tyto údaje nemají všechny databáze, bylo rozhodnuto, že se kraj bude



Obr. 4.4: Poměr lokalizace.



Obr. 4.5: Kumulativní distribuční funkce pro lokační databáze.

určovat podle příslušnosti daného města nebo lokalizovaného místa. Vyhodnocení přesnosti města probíhalo tak, že byly vzaty v úvahu pouze údaje v položce město. To v praxi znamená, že se lokalizovaný bod pomocí souřadnic může nacházet několik kilometrů za hranicí města, ale na výsledek to nemá žádný vliv.

Detekce kraje bude probíhat odlišně a to tak, že budou vzaty souřadnice bodu a budou přeloženy na kraj, kde se bod nachází. Je to zejména z důvodu, že by vypadla databáze IP Info, která v mnoha případech měla v položce null. Z pohledu ovlivnění přesnosti výsledků není tento postup jako velký problém, protože jednotlivé kraje na sebe navazují a není prostor pro vznik zkreslení. Lokalizovaný bod je buď členem jednoho kraje nebo jiného. Totéž nelze tvrdit o lokalizaci na úrovni měst nebo vesnic. Jednotlivá města nebo vesnice na sebe nenavazují a v případech, kdy se lokalizovaný bod nachází mezi dvěma městy, mohou nastat různé výsledky do jakého města nebo vesnice lokalizovaný bod zařadí.

Posledním z vyhodnocovaných výsledků bude přesnost na základě státu. Tyto výsledky jsou v rámci geolokace velmi nepřesné, ale v mnoha případech, kdy je nutné ověřit polohu uživatele například pro přístup k internetovému vysílání televizních stanic, je tato přesnost naprosto dostatečná. Omezení přístupu k online přenosu televizního signálu pro zahraniční návštěvníky má na stránkách Česká televize.

Zpracování výsledků probíhá ručně a provádí se kontrola, zda vložené výsledky odpovídají výsledkům detekovaným. V případě, kdy dojde ke správné lokalizaci, je proveden zápis do dat ke zpracování jako ano. V opačném případě bude výsledek označen jako neplatný.

Detekce měst a obcí

Princip detekce města a obce je popsán o pár odstavců výše, proto budou hned vzaty v potaz naměřené výsledky. V tabulce 4.5 jsou uvedeny v procentech správně detekována města a obce. Nebere se ohled na to, zda výsledek nebyl nalezen v databázi nebo výsledek byl dříve označen jako nelokalizovatelný, tedy jeho odchylka od správné lokace byla větší než 500 km. Jistou výhodou mají IP adresy z velkých měst, hlavně tedy z krajských a bývalých okresních měst. Důvod, proč tomu tak je lze vysvětlit tím, že lokální poskytovatelé působící na menších oblastech mají právě sídla v krajských nebo v bývalých okresních městech. Přitom často pokrývají svými službami i okolní obce. U větších poskytovatelů DSL připojení je často jako místo připojení detekována Praha. Z toho vyplývá, že IP adresy z Prahy mají větší pravděpodobnost správné lokace a nelze zjistit, zda to byl výsledek naměřený nebo si s IP adresou databáze nedokázala poradit a přiřadila Prahu, protože již velké množství IP adres z daného rozsahu se nachází v Praze. Je také důležité zmínit, že detekovat obce je téměř nemožné. Nebylo provedeno žádné měření, jak velikou obec již

lze téměř bez problému detekovat. Naopak z mapy historie naměřených výsledků je možné vyčíst, že města s velikostí pohybující se okolo 4 000 obyvatel, lze s velkou pravděpodobností rozpoznat.

Z naměřených výsledků nejlépe detekovala IP adresy databáze Skyhook, se svými 43 % má opět velký náskok před konkurenčními databázemi. Databáze DB-IP.com dokázala detekovat IP adresy s druhou nejvyšší přesností 36 %, Maxmind Lite a IP Info předvedli zcela schodný výsledek 33 %. Nejhorší výsledek znovu předvedla databáze Geobytes, která naměřila nejhorší výsledky ze všech doposud vyhodnocovaných testů.

Skyhook Wireless	DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
43 %	36 %	19 %	33 %	33 %

Tab. 4.5: Úspěšnost detekce města nebo obce.

Detekce na úrovni krajů / regionů

Přesnost na úrovni krajů není, co se týče požadavků na lokační databáze, tak náročná, přesto dokáže přinést zajímavé výsledky vhodné zejména pro regionální cílení na uživatele, například u rádií. Při detekci přístupu na web bude tato přesnost dostatečná. Další oblastí, kde dosahovaná přesnost může být dostatečná, je předpověď počasí. Meteorologické mapy mají rozlišení v řádech kilometrů, nabízení automatické předpovědi pro detekovaný kraj může být pro většinu návštěvníků dostatečné.

Databáze z pohledu detekce krajů nabídl tři výsledky. V nejlepším případě bylo měření úspěšné. V druhém případě došlo k detekci sousedního kraje a to obvykle tehdy, kdy se lokalizovaná IP adresa nacházela v okolí hranice kraje. Je otázkou diskuze, zda měření považovat za nesprávné z pohledu uživatele případě, kdy se například jedná o zmiňovanou předpověď počasí. Ve třetím případě byl detekován zcela odlišný kraj a takový výsledek, lze bez pochyby označit jako neplatný. Výsledky byly vyhodnoceny tak, že pokud se neshodoval kraj, ve kterém se nacházela IP adresa s detekovaným krajem byl výsledek považován za neplatný i v případě, kdy se IP adresa nacházela poblíž hranic krajů.

Z naměřených výsledků v tabulce 4.6, lze pozorovat, že tři databáze se srovnali na rozmezí jednoho procenta přesnosti, což odpovídá jedné rozdílně detekované IP adrese. Nejvíce v přesnosti rostla databáze Skyhook Wireless a to na 72 %. Vezmeme-li v úvahu, že databáze dokáže maximálně detekovat 89 % IP adres, je přesnost na úrovni detekci krajů velmi na vysoké úrovni. Databázi Geobytes.com nepomohlo ani rozšíření prostoru pro detekci i tak má úspěšnost detekce 37 %. Lze tedy říci, že databázi Geobytes.com nelze použít pro geolokaci na úrovni krajů v České republice.

Skyhook Wireless	DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
72 %	53 %	37 %	53 %	52 %

Tab. 4.6: Úspěšnost detekce kraje.

Detekce státu

Detekci státu lze považovat za základní stavební jednotku správné geolokace. Pokud není geolokační databáze ve většině případů schopna správně detekovat alespoň stát, je použitelnost takové databáze téměř nulová. Na správné detekci státu lze postavit automatické překlady do jazyka stránek nebo omezení přístupu na stránky, které nabízejí obsah pouze pro občany dané země. Pro vývojáře aplikací, kteří by se chystali stavět stránky na správné detekci, je správná detekce státu kritická. Situace na webu již zmiňované České televize, která povoluje získání obsahu archívu pouze pro obyvatele České republiky, by při použití databáze Geobytes vypadala tak, že by téměř pro 25 % návštěvníků by byla služba zablokována.

Výsledky správnosti detekce na úrovni státu jsou uvedeny v tabulce 4.7. Téměř všechny databáze dosáhli úspěšnosti nad 90 % opět s výjimkou databáze Geobytes.com, která má evidentně velmi nekvalitní databázi výsledků. Další databáze dosáhli 100 % a to lze považovat za velmi kvalitní základ. Databáze Skyhook Wireless nedosáhla přesnosti 100 % to proto, že některé IP adresy nebyla schopna správně detekovat.

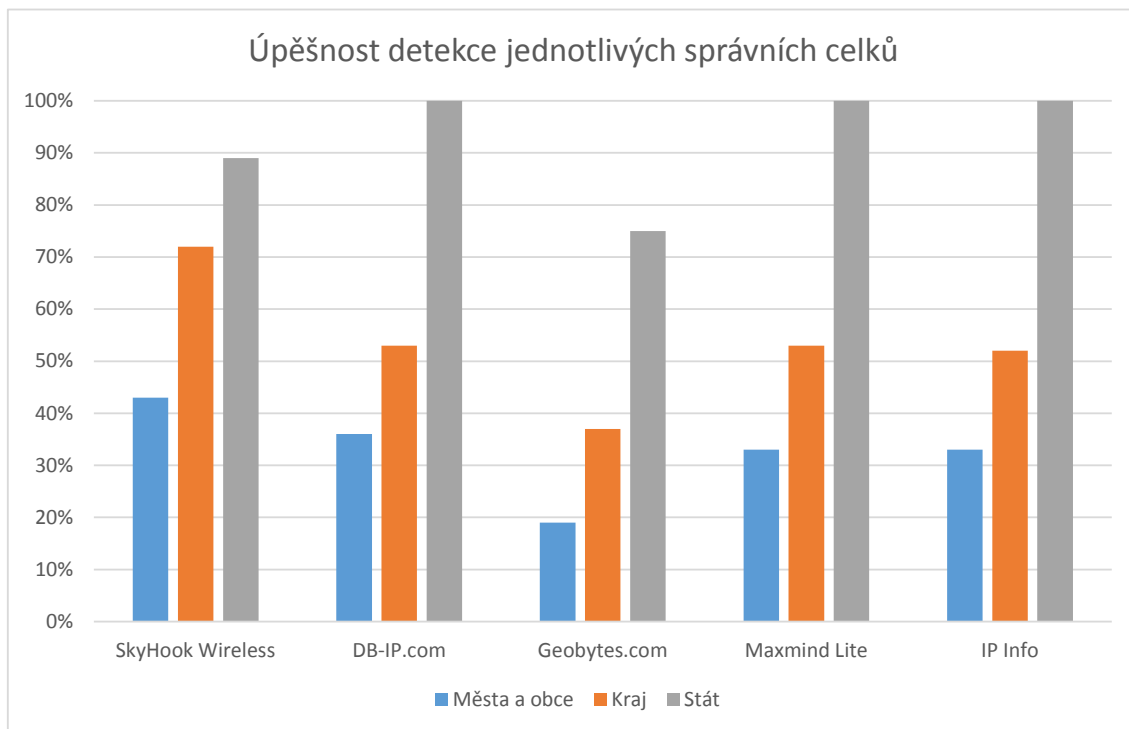
Na obrázku 4.6 je graf shrnující všechna data naměřená a zpracována na základě správních celků a přehledně tak shrnuje data uvedená v předchozích třech podkapitolách.

Skyhook Wireless	DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
89 %	99 %	75 %	100 %	100 %

Tab. 4.7: Úspěšnost detekce na úrovni státu.

4.1.5 Zhodnocení výsledků

Jelikož nelze vzít jeden naměřený výsledek a označit nejpřesnější databázi, byly stanoveny následující pravidla. Všechna měření mají přidělenou váhu a pro každou skupinu je určeno pořadí výsledků. Skupina, na jejich základě probíhá dělení výsledků, je následující: průměrná naměřená odchylka databází, podle procentuálního rozložení naměřených vzdáleností a přesnost lokačních databází na základě správních celků. Skupina procentuálního rozložení výsledků je dále rozdělena na kategorie



Obr. 4.6: Graf úspěšnosti detekce na základě správních celků.

odchylek do 10 km, do 50 km, do 100 km, do 250 a celkem. Skupina rozpoznání správních celků je dále dělena do kategorií na rozpoznání města nebo obce, kraje a státu. Maximálně bude možné dosáhnout hranice 90 bodů. Bodové hodnocení je uvedené v tabulce 4.8

Pořadí	Body
1	10
2	7
3	5
4	3
5	1

Tab. 4.8: Bodové hodnocení výsledků.

Body přidělené pro skupinu průměrné naměřené odchylky jsou uvedeny v tabulce 4.9. Maximální bodový zisk je 50 bodů. Pro přehlednost jsou zde uvedeny pouze body celkem za skupinu tak, aby vyhodnocení výsledků zůstalo pokud možno co nejvíce přehledné. Jednotlivé kategorie lze dohledat v příložených grafech.

Jak je vidět, počet bodů na prvních třech místech je téměř vyrovnaný. Podle výsledků vypadalo, že nejlepší musí být Skyhook Wireless, ale právně vysoký počet

Skyhook Wireless	DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
36	25	5	35	35

Tab. 4.9: Bodový zisk ve skupině procentuální rozložení naměřených odchylek.

nelokalizovaných adres dává šanci i ostatním databázím a to nemuselo být na první pohled zřejmé, proto jsou zde uvedeny body.

Body ve skupině rozpoznání správních celků jsou uvedeny v tabulce 4.10. Maximální bodový zisk činí 30 bodů. Největší počet bodů získala databáze Skyhook Wireless, ale opět je pořadí na prvních třech místech v rozmezí čtyř bodů. Nejnižší počet bodů má databáze Geobytes.com, která byla v každé kategorii na posledním místě.

Skyhook Wireless	DB-IP.com	Geobytes.com	Maxmind Lite	IP Info
25	21	3	22	20

Tab. 4.10: Bodový zisk ve skupině rozpoznání správních celků.

Celkový bodový zisk je uveden v tabulce 4.11 a je seřazen podle pořadí celkových výsledků. Tabulka shrnuje všechny skupiny včetně průměrné naměřené odchylky, která nemá samostatnou tabulku, protože by došlo k duplicitě s dříve uvedenou tabulkou. Maximální počet bodů, které bylo možné za všechny skupiny získat je 90 bodů.

Skyhook Wireless	Maxmind Lite	IP-Info	DB-IP	Geobytes.com
71	64	60	49	8

Tab. 4.11: Bodový zisk celkem i s body z průměrné naměřené odchylky.

Nejvyšší počet bodů získala databáze Skyhook Wireless, která již podle naměřených výsledků pracovala s velmi přesnou databází. Překvapením je ne příliš velký rozdíl mezi prvními třemi databázemi. Předběžné výsledky naznačovaly jasnou převahu databáze Skyhook Wireless, ale právě vysoký počet nelokalizovaných IP adres databázi Skyhook Wireless připravilo o jasnou převahu. Nelokalizované IP adresy mohou v reálném nasazení znamenat značné komplikace, a proto v případě, kdy by byla použita databáze Skyhook Wireless by bylo vhodné také použít druhou lokační databázi a to právě druhou v pořadí databázi Maxmind Lite a nebo z rozboru uvedenou databázi DB-IP.com. Je nutné na závěr uvést, že výsledky jsou platné především pro Českou republiku. Pro jiné země, výsledky mohou dosahovat značně odlišných hodnot a například databáze Geobytes.com může na území Severní Ameriky podávat značně odlišné výsledky.

5 ZÁVĚR

Zadáním diplomové práce bylo vytvořit webový systém, který bude umožňovat zobrazení komplexních informací o zadané nebo detekované IP adrese. V první kapitole je definováno, co je geolokace v síti internet a jaké jsou důvody, proč se geolokace využívá. Také je popsáno, jaké služby jsou nebo mohou být postaveny na geolokaci a co to znamená pro uživatele na internetu.

Ve druhé kapitole jsou blíže přiblíženy metody geolokace, které se běžně používají především v mobilních zařízeních. Jedná se o geolokaci pomocí základnových stanic (BTS), pomocí družicového signálu GPS a určení polohy pomocí Wi-Fi sítě. Dále se kapitola věnuje tématu geolokace v sítích IP a vznikajících problémech v rozsáhlých sítích využívající NAT a dále pak přibližuje možnosti jakými je možné získat údaje o poloze IP adresy. Jsou zde zmíněny jak aktivní geolokační metody, tak i pasivní geolokační metody, které jsou pak blíže specifikovány. Sběr dat do geolokačních databází patří mezi nejkomplicovanější a také patří mezi nejvíce chráněnou informaci ve firemním sektoru. Byl zde proto popsán princip, který může být použit pro prvotní naplnění geolokačních databází a tyto data pak mohou být dále zpřesňována. Jedná se o databázi přidělených IP adres organizací IANA.

Geolokační databáze jsou nejvíce efektivním způsobem, jak zjistit polohu jakékoli veřejné IP adresy v síti internet. Databáze lze rozdělit na placené a na databáze dostupné zdarma. Diplomová práce se zabývá zkoumáním výsledků placených geolokačních databází. Jedna databáze je dostupná offline a další čtyři jsou umístěny na jiných serverech v internetu. Práce také zvažuje, zda je vhodnější použít databázi dostupnou online nebo využít offline databázi. Dále je v kapitole popsána struktura geolokační databáze, a ke každé použité databázi ve webovém systému jsou uvedeny základní informace.

Ve třetí kapitole je popsán zhotovený systém pro geolokaci stanic v síti internet. Systém využívá pět placených databází, pro které jsou napsány funkce přímo ve webovém systému. Další data z geolokačních databází jsou získávána již v definovaném formátu od kolegů, kteří pracovali s volně dostupnými geolokačními databázemi a s databází Whois. Jako doplňující informace jsou označeny informace o detekci proxy serveru a o měření zpoždění. V následujících podkapitolách je popsáno, jakým způsobem jsou získávány vstupní informace a jakým způsobem jsou generovány dotazy na geolokační databáze. Každá databáze má odlišný formát dotazu a odlišný formát odpovědi a pořadí dat, proto jsou data nejdříve převedena na společný formát, aby bylo možné co nejjednodušeji zpětně vyčítat. Výsledky jsou prezentovány do tabulky a pro názornost jsou zobrazeny v mapě. Pokud je známá správná lokace IP adresy, která může být získána překladem z adresy na souřadnice. Tato funkce je umožněna API od společnosti Google. Vzdálenost mezi body je

vypočítána z detekovaných a převedených souřadnic pomocí Vincencova algoritmu. Výsledkem je vzdálenost v metrech. Vincencův algoritmus patří mezi nejpřesnější algoritmy využívající se pro přepočítání souřadnic na vzdálenost mezi body. K ukládání výsledků lokalizace IP adres dochází, pouze pokud jsou známy správné souřadnice IP adresy. Za správné souřadnice se považuje každá úspěšně převedená adresa nebo přímo uživatelem vložené souřadnice. Žádné další ověřování adres nebo souřadnic není prováděno. Data jsou ukládána do textového souboru, jeho struktura a funkce je popsána v kapitole 3.1.6. Dalším rozšířením webového systému je zobrazení uložených výsledků přímo na webu. Výsledky jsou prezentovány tak, jako jsou zobrazeny uživateli. Pro názornost jsou také zobrazeny do mapy.

Ve čtvrté kapitole je diskutována a vyhodnocena přesnost geolokačních databází. Vyhodnocení výsledků je rozděleno na dvě skupiny. První skupina tvoří vyhodnocení výsledků podle vzdálenosti. Nejdříve je z naměřených dat vypočítána průměrná odchylka pro jednotlivé databáze. Dále jsou výsledky rozděleny do skupin podle chyby lokace. Každá skupina je vyhodnocena zvlášť. Další kategorií je efektivita lokace, která je velmi rozdílná a může se stát rozhodujícím faktorem při volbě geolokační databáze. Dalším testovaným parametrem je správné rozpoznání správních celků. V České republice se jedná o detekci města, kraje a státu. V poslední kapitole 4.1.5 jsou zhodnoceny výsledky. Jednotlivé skupiny výsledků jsou obodovány a podle celkových bodů je zvoleno pořadí databází.

LITERATURA

- [1] LEIBA, B. 2013. Update to Internet Message Format to Allow Group Syntax in the "From:"and "Sender:"Header Fields. *Request for comments published by the Internet Engineering Task Force* [online] (Request for Comments: 6854). [cit. 10. 5. 2015]. ISSN 2070-1721. Dostupné z URL: <<https://tools.ietf.org/html/rfc6854>>.
- [2] KORCH Robert, Mario GOLLING a Dreo RODOSEK. *Advanced Geolocation of IP Addresses* [online]. 2013, [cit. 7. 10. 2014]. Dostupné z URL: <<http://waset.org/publications/16111/advanced-geolocation-of-ip-addresses>>.
- [3] KOMOSNÝ Dan, VERNER Lukáš. *Geolokace síťových zařízení v internetových sítích* Elektrotechnika. 2011, [cit. 7. 10. 2014].
- [4] AudioCodes. *Geographical location information for emergency calls over VoIP networks* [online]. 2009, [cit. 7. 10. 2014]. Dostupné z URL: <<http://www.audiocodes.com/filehandler.ashx?fileid=452775>>.
- [5] *Základní lokalizační metody v GSM* [online]. 2006, [cit. 8. 10. 2014]. Dostupné z URL: <<http://access.feld.cvut.cz/view.php?cisloclanku=2006022801>>.
- [6] *The Global Positioning system: A Shared National Asset: Recommendations for Technical Improvements and Enhancements.* [online]. Washington, D.C.: National Academy Press, 1995, xix, 264 p. ISBN 03-090-5283-1. Dostupné z URL: <http://www.nap.edu/catalog.php?record_id=4920>.
- [7] VAUGHAN-NICHOLS, Steven J. ZDNet. *How Google-and everyone else-gets Wi-Fi location data.* [online]. 2011 [cit. 2014-11-19]. Dostupné z URL: <<http://www.zdnet.com/blog/networking/how-google-and-everyone-else-gets-wi-fi-location-data/1664>>.
- [8] RIPE NCC RIPE NCC. *Annual Report* [online]. 2006 [cit. 2014-11-25]. Dostupné z URL: <http://www.ripe.net/lir-services/ncc/gm/may-2007/supporting-documents/annual_report_2006.pdf>.
- [9] IANA IPv4 Address Space Registry. IANA. *Internet Assigned Numbers Authority (IANA)* [online]. 2015 [cit. 2015-05-02]. Dostupné z URL: <<http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xhtml>>.

- [10] P geolocation database downloads: MySQL scripts *DB-IP - IP Geolocation and Network Intelligence* [online]. 2014 [cit. 2015-05-03]. Dostupné z URL: <<https://db-ip.com/db/#mysql>>.
- [11] HYPERLOCAL IP. *Skyhook / Global Location and Context Software Products*. [online]. 2014 [cit. 2014-11-19]. Dostupné z URL: <<https://my.skyhookwireless.com/resources/view/hyperlocalip/api>>.
- [12] IP Geolocation online API. *DB-IP - Your ultimate resource for IP geolocation*. [online]. 2014 [cit. 2014-11-19]. Dostupné z URL: <<https://db-ip.com/api/>>.
- GeobytesAPICite
- [13] IP Address Locator. *Geobytes* [online]. 29.08 2006 [cit. 2014-11-21]. Dostupné z URL: <<http://dev.maxmind.com/geoip/geoip2/geolite2/>>.
- [14] GeoLite2 Free Downloadable Databases. *Maxmind Developer Site* [online]. 2014 [cit. 2014-11-21]. Dostupné z URL: <<http://www.geobytes.com/IpLocator.htm?GetLocation>>.
- [15] Developers. *IP Address Details - ipinfo.io* [online]. 2014 [cit. 2014-11-23]. Dostupné z URL: <<https://ipinfo.io/developers>>.
- [16] THOMAS, C. M. a W. E. FEATHERSTONE. Validation of Vincenty's Formulas for the Geodesic Using a New Fourth-Order Extension of Kivioja's Formula. In: *Journal of Surveying Engineering* [online]. 2015, s. 20–26 [cit. 2014-12-08]. ISSN 0733-9453. DOI: 10.1061/(ASCE)0733-9453(2005)131:1(20). Dostupné z URL: <[http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)0733-9453\(2005\)131:1\(20\)](http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9453(2005)131:1(20))>.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

AfriNIC African Network Information Center

API Application Programming Interface

APNIC Asia Pacific Network Information Centre

ARIN American Registry for Internet Numbers

BTS Base Transceiver Station

CSS Cascading Style Sheets

DNS Domain Name System

GPS Global Positioning System

HTML HyperText Markup Language

IANA Internet Assigned Numbers Authority

ICANN Internet Corporation for Assigned Names and Numbers

HTTP Hypertext Transfer Protocol

HTTPS Hypertext Transfer Protocol Secure

IP Internet Protocol

ISP Internet Service Provider

JSON JavaScript Object Notation

LACNIC Latin America and Caribbean Network Information Centre

MAC Media Access Control

PHP Hypertext Preprocessor

RIPE NCC Réseaux IP Européens Network Coordination Centre

RIR Regional Internet registry

SSID Service Set Identifier

URI Uniform Resource Identifier

VoIP Voice over Internet Protocol

XML eXtensible Markup Language

A OBSAH PŘILOŽENÉHO CD

- Elektronická verze diplomové práce.
- Zdrojové kódy webové aplikace.