

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

Evidenčné číslo: FEI-5382-52598

**VYUŽITIE GRAFOVEJ DATABÁZY V PRAXI
BAKALÁRSKA PRÁCA**

2017

Juraj Kubričan

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Evidenčné číslo: FEI-5382-52598

VYUŽITIE GRAFOVEJ DATABÁZY V PRAXI
BAKALÁRSKA PRÁCA

Študijný program: Aplikovaná informatika
Číslo študijného odboru: 2511
Názov študijného odboru: 9.2.9 Aplikovaná informatika
Školiace pracovisko: Ústav informatiky a matematiky
Vedúci záverečnej práce: Ing. Maroš Čavojský

Bratislava 2017

Juraj Kubričan



ZADANIE BAKALÁRSKEJ PRÁCE

Študent: **Juraj Kubričan**
ID študenta: **52598**
Študijný program: **aplikovaná informatika**
Študijný odbor: **9.2.9. aplikovaná informatika**
Vedúci práce: **Ing. Maroš Čavojský**
Miesto vypracovania: **Ústav informatiky a matematiky**

Názov práce: **Využitie grafovej databázy v praxi**

Jazyk, v ktorom sa práca vypracuje: **slovenský jazyk**

Špecifikácia zadania:

V dnešnej dobe sa okrem tradičných zaužívaných relačných databáz, využívajú aj menej známe grafové databázy, v ktorých sú dáta uložené odlišným spôsobom ako v relačných databázach. Cieľom práce je oboznámiť sa s jednotlivými predstaviteľmi grafových databáz, vybrať jedného a navrhnúť a implementovať využitie vybranej grafovej databázy na reálnom príklade.

Úlohy:

1. Naštudujte si literatúru ohľadom jednotlivých predstaviteľov grafových databáz
2. Vyberte jedného predstaviteľa grafových databáz
3. Navrhnite reálny príklad pre implementáciu grafovej databázy
4. Implementujte reálny príklad pre implementáciu grafovej databázy
5. Zhodnoťte a uveďte výhody použitia grafovej databázy oproti iným typom databáz (relačné, dokumentové,...) v implementovanom reálnom príklade

Zoznam odbornej literatúry:


1. Ian Robinson, Jim Webber, and Emil Eifrem: Graph Databases, O'Reilly Media, Inc. USA 2015, p.224, ISBN: 978-1-491-93200-1


Riešenie zadania práce od: **19. 09. 2016**

Dátum odovzdania práce: **19. 05. 2017**


Juraj Kubričan
študent

SLOVENSKÁ TECHNICKÁ UNIVERZITA
V BRATISLAVE
Fakulta elektrotechniky a informatiky
Ústav informatiky a matematiky
Ilkovičova 3, 812 19 Bratislava


prof. RNDr. Otokar Grošek, PhD.
vedúci pracoviska


prof. Dr. Ing. Miloš Oravec
garant študijného programu

SÚHRN

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

| | |
|---------------------------------|------------------------------------|
| Študijný program: | Aplikovaná informatika |
| Autor: | Juraj Kubričan |
| Bakalárska práca: | Využitie grafovej databázy v praxi |
| Vedúci záverečnej práce: | Ing. Maroš Čavojský |
| Miesto a rok predloženia práce: | Bratislava 2017 |

Práca sa zaoberá vytvorením webovej aplikácie ktorá bude využívať grafovú databázu. V prvej časti sa nachádza prehľad technológií, ktoré sme použili na implementáciu projektu, ďalej sa tu nachádza priblíženie najpopulárnejších zastupiteľov grafových databáz, ich výhod a nevýhod. V ďalšej časti sa nachádza špecifikácia našej aplikácie cestovného plánovača.

Kľúčové slová: Využitie grafovej databázy v praxi

ABSTRACT

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA

FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

| | |
|-------------------------------|--|
| Study Programme: | Applied Informatics |
| Author: | Juraj Kubričan |
| Bachelor Thesis: | Graph database use in a real world application |
| Supervisor: | Ing. Maroš Čavojský |
| Place and year of submission: | Bratislava 2017 |

The bachelor thesis is about creating of a plugin for web browser, that modifies information used to identification of user during accessing a server. There is an overview of methods that increase anonymity during browsing websites, in the first part. The thesis also contains a list of the most used extensions nowadays, that function is a change of some identification components of browser or special ways of anonymization. In the next part of the thesis is an overview of the characteristics of web browser. By combination of these characteristics we can with high level of success identify a user, who have visited the web site. The last part of thesis contains project, implementation and testing of extension created for the web browser Mozilla Firefox. There is also description of source code of extension, the link between the characteristics of web browser, detected limitations and way how to solve them. The resulting extension increases anonymity of user by modification of some characteristic components of web browser or by blocking sending components, that can not be in extension changed. In comparison with most used modules nowadays, this module can modify HTTP headers including characteristics detected by JavaScript commands.

Keywords: Graph database use in a real world application

Obsah

| | |
|---|-----------|
| Úvod | 1 |
| 1 Analýza problému | 2 |
| 1.1 Relačné | 2 |
| 1.2 Grafové | 2 |
| 2 Cieľ práce | 3 |
| 3 Proces tvroby systému | 4 |
| 4 Špecifikácia | 5 |
| 5 Vývoj | 7 |
| 5.1 Návrh | 7 |
| 5.1.1 Usecase diagramy | 7 |
| 5.1.2 Class diagram | 7 |
| 5.1.3 Role hráčov | 7 |
| 5.2 Implementácia | 10 |
| 5.2.1 Výber grafovej databázy | 10 |
| 6 Implementácia | 12 |
| 6.1 Použíte technológie | 12 |
| 6.1.1 Laravel framework | 12 |
| 6.1.2 NeoEloquent OGM | 12 |
| 6.1.3 Mapbox.js | 12 |
| 6.1.4 Rome2Rio Api | 12 |
| 6.2 Vzhľad | 12 |
| Záver | 14 |
| Zoznam použitej literatúry | 15 |
| Prílohy | I |

Zoznam obrázkov a tabuliek

| | | |
|-----------|--|----|
| Obrázok 1 | Usecase - Registrácia, autentikácia a nastavenia | 8 |
| Obrázok 2 | Usecase - dashboard | 9 |
| Obrázok 3 | Usecase - tsp | 9 |
| Obrázok 4 | Class diagram | 10 |

Zoznam skratiek a značiek

TSP - The Travelling Salesman Problem

UML - Unified Modeling Language verzie 2.5

GDBMS RDBMS GPLv3

Zoznam algoritmov

Úvod

Pri návrhu aplikácie treba myslieť na štruktúru dát a podľa toho vybrať správnu databázu/DBMS. Správny výber DBMS vie zabezpečiť rádovo nižšie prístupové časy a tým aj väčšiu scalability. Pri aplikáciách kde sú dáta štruktúrované do uzlov a prepojení, je vhodné zvážiť použitie grafovej databázy. My sme navrhli a naimplementovali aplikáciu cestovného plánovača, ktorý potrebuje uchovávať dáta o destináciách a cestách medzi nimi. Preto je pre túto aplikáciu vhodné využiť grafový DBMS

1 Analýza problému

1.1 Relačné

Relačná databáza je databáza, v ktorej sú údaje uložené podľa relačného databázového modelu podľa E. F. Codda z roku 1970. Podľa Relačného modelu sú dáta uložené v tabuľkách. Jeden riadok tabuľky je jeden záznam. Stĺpec tabuľky reprezentuje jeden atribút objektu. Väzby (vzťahy) medzi tabuľkami sú riešené pomocou unikátnych identifikátorov tzv. kľúčov. Jheda tabuľka obsahuje kľúč, čo je atribút ktorý unikátne identifikuje každý záznam a druhá tabuľka obsahuje tzv cudzí kľúč, atribút ktorý odkazuje na záznam v prvej tabuľke. Výhodou tohto spôsobu ukladania dát je jednoduchosť a prehľadnosť. Nevýhoda tohto prístupu sa však ukazuje v škálovateľnosti. Pri vyhľadávaní každé toto prepojenie pridáva výpočtovú komplexitu, keďže v každej ďalšej prepojennej tabuľke treba vyhľadať záznam s požadovaným kľúčom ($O(\log(n))$). Všetky používané relačné databázové systémy riešia tento problém škálovateľnosti použitím indexov a rôznymi inými optimalizáciami, no pokiaľ sú naše dáta štrukturované s veľa prepojeniami systém sa spomalí.

1.2 Grafové

V grafovej databáze sú údaje štrukturované vo vrchoch(node) a hranách(edge). Prepojenia medzi vrcholmi sú realizované priamo pomocou hrán. Grafové databázové systémy podporujú vlastné sémantické prostriedky ktoré umožňujú priamu manipuláciu so štruktúrou dát. Jedna z hlavných výhod Táto štruktúra umožňuje priamo

2 Ciel' práce

Cielom práce je navrhnuť a na implementovať aplikáciu, ktorá využíva grafovú databázu, ďalej zhodnotiť výhody tohto riešenia oproti iným typom databáz.

3 Proces tvorby systému

4 Špecifikácia

1. Funkcionálne požiadavky

- (a) Aplikácia bude umožňovať registráciu a prihlásenie používateľa
- (b) Pri registrácii sa budú vyžadovať prihlasovacie údaje: e-mail, heslo. Okrem toho sa bude vyžadovať zadanie mena a domáceho miesta. Toto domáce miesto bude možné vyhľadať v online databáze.
- (c) Po prihlásení používateľa sa mu zobrazí obrazovka s mapou, zoznamom obľúbených destinácií, ktoré chce navštíviť a zoznam odporúčaných destinácií
- (d) Na mape bude vyobrazené používateľovo domáce miesto a všetky destinácie, ktoré má v zozname obľúbených destinácií. Po kliknutí na destináciu sa používateľovi otvorí príslušný riadok v zozname obľúbených.
- (e) V zozname obľúbených budú všetky miesta, ktoré si používateľ pridal. Zoznam bude vo forme tabuľky, riadok bude obsahovať meno destinácie a lokalitu, v ktorej sa destinácia nachádza. V riadku tiež bude tlačidlo na vymazanie destinácie z obľúbených a tlačidlo na zobrazenie detailov.
- (f) V detailoch obľúbeného miesta bude zoznam ostatných ľudí, ktorí dané miesto majú v obľúbených a výpis možných trás z domáceho miesta používateľa do destinácie.
- (g) V zozname odporúčaných destinácií budú destinácie, ktoré majú v obľúbených používatelia, ktorí majú v obľúbených rovnaké miesta ako miesta, ktoré má v obľúbených prihlásený používateľ. Vynechané budú miesta, ktoré už prihlásený používateľ má v obľúbených. Každá položka z odporúčaných sa bude dať jednoducho pridať do obľúbených prihláseného používateľa.
- (h) V aplikácii bude obrazovka s nastaveniami, na ktorej si bude človek môcť zmeniť heslo, domáce miesto.
- (i) V aplikácii bude obrazovka, kde si bude používateľ vybrať niekoľko zo svojich obľúbených miest a nechať si vyrátať najkratšiu trasu z domáceho miesta cez všetky zvolené miesta a potom späť. (TSP)

2. Nefunkcionálne požiadavky

- (a) Systém bude zrealizovaný na webovej platforme.
- (b) Aplikácia bude využívať natívnu grafickú databázu.

- (c) Aplikácia bude byť kompatibilná s webovými prehliadačmi Google Chrome Mozilla Firefox, Microsoft Edge, Microsoft Internet Explorer.
- (d) Užívateľské rozhranie systém musí byť plne použiteľné aj na mobilných telefónoch s OS Android a IOS.
- (e) Aplikácia bude implementovaný s použitím jayka PHP a PHP frameworku.
- (f) Systém bude nasadený na virtuálnom serveri s operačným systémom Ubuntu 16.04.2 LTS poskytnutom Ústavom informatiky a matematiky FEI STU.

5 Vývoj

5.1 Návrh

V časti návrhu projektu popíšeme prípady použitia,

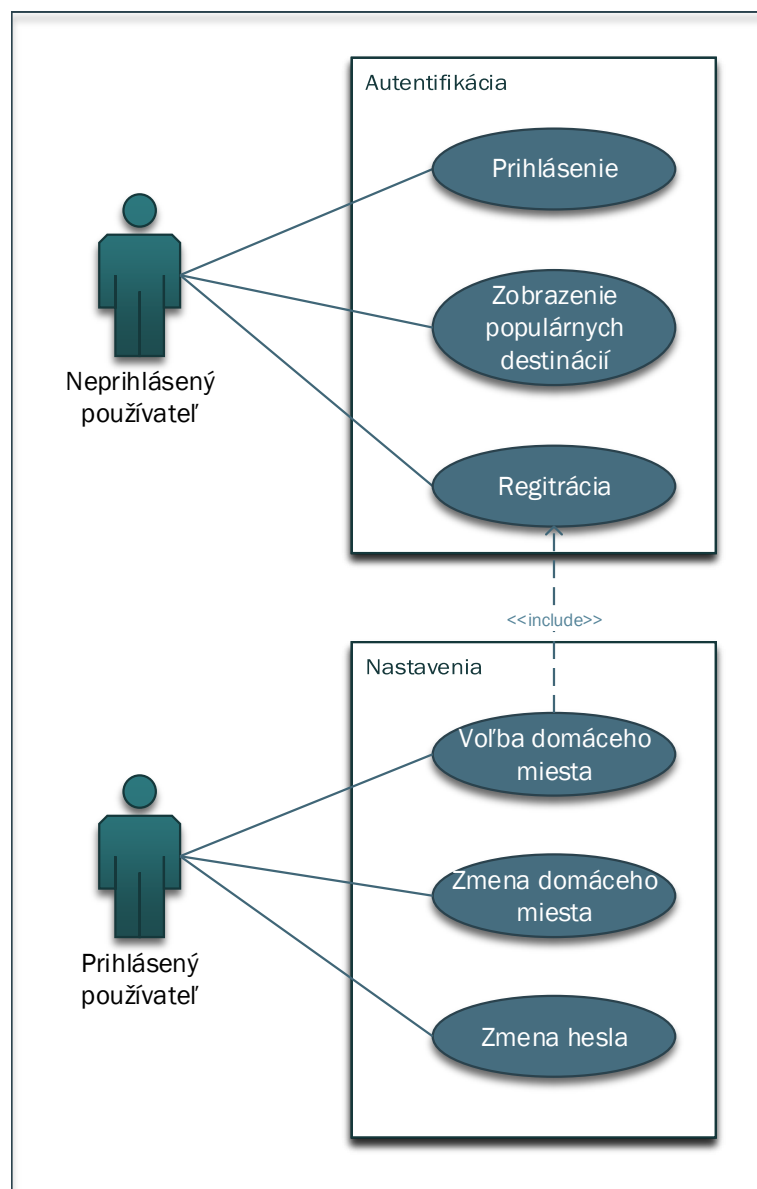
5.1.1 Usecase diagramy

1. Usecase - Registrácia, autentikácia a nastavenia
2. Usecase - dashboard
3. Usecase - ts

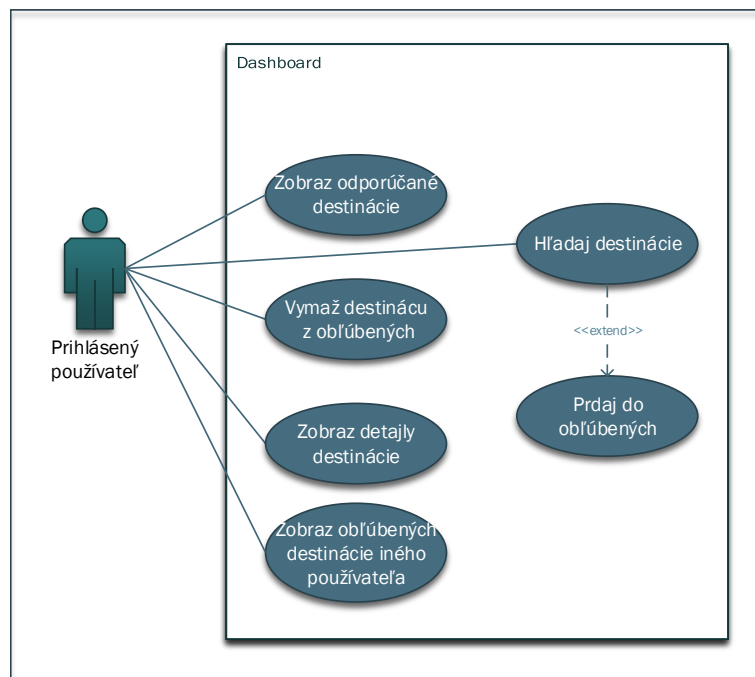
5.1.2 Class diagram

5.1.3 Role hráčov

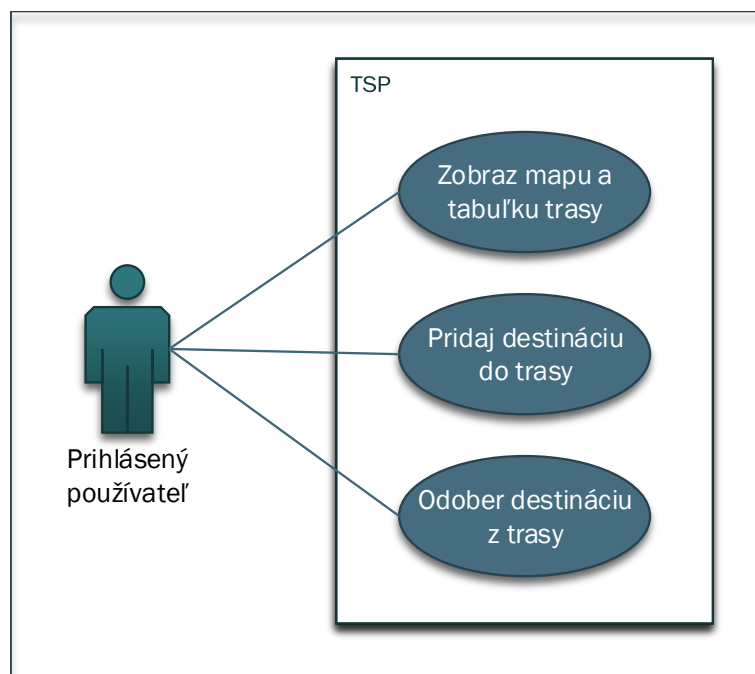
1. Neprihlásený používateľ bude mať prístup na úvodnú stránku. Na úvodnej stránke sa bude môcť buď zaregistrovať, prihlásiť ak už má vytvorený účet, alebo si bude môcť pozrieť mapu a zoznam s najpopulárnejšími destináciami používateľov aplikácie.
2. Prihlásený používateľ má prístup do štyroch subsystémov
 - (a) Dashboard je domáca obrazovka používateľa bude môcť na nej vyhľadávať, pridávať a odstraňovať miesta medzi svoje obľúbené. Ďalej si bude môcť pozrieť detaily destinácie ako rôzne trasy ktoré vedú z jeho domáceho miesta do destinácie a zoznam ostatných používateľov, ktorí majú rovnaké miesta v obľúbených. Kliknutím na používateľa bude môcť prejsť na obrazovku používateľa.
 - (b) Na obrazovke nastavení si bude môcť používateľ nastaviť heslo a zmeniť svoje domáce miesto.
 - (c) Na obrazovke TSP si bude používateľ môcť naplánovať trasu medzi niektorými zo svojich obľúbených miest. Bude si môcť pridávať miesta do trasy a odoberať ich. Ďalej si bude môcť prezrieť detaily aktuálnej trasy s cenami a mapu s vyobrazenou trasou.
 - (d) Na obrazovke používateľa si bude prihlásený používateľ prezrieť obľúbené destinácie konkrétneho používateľa na mape a v zozname.



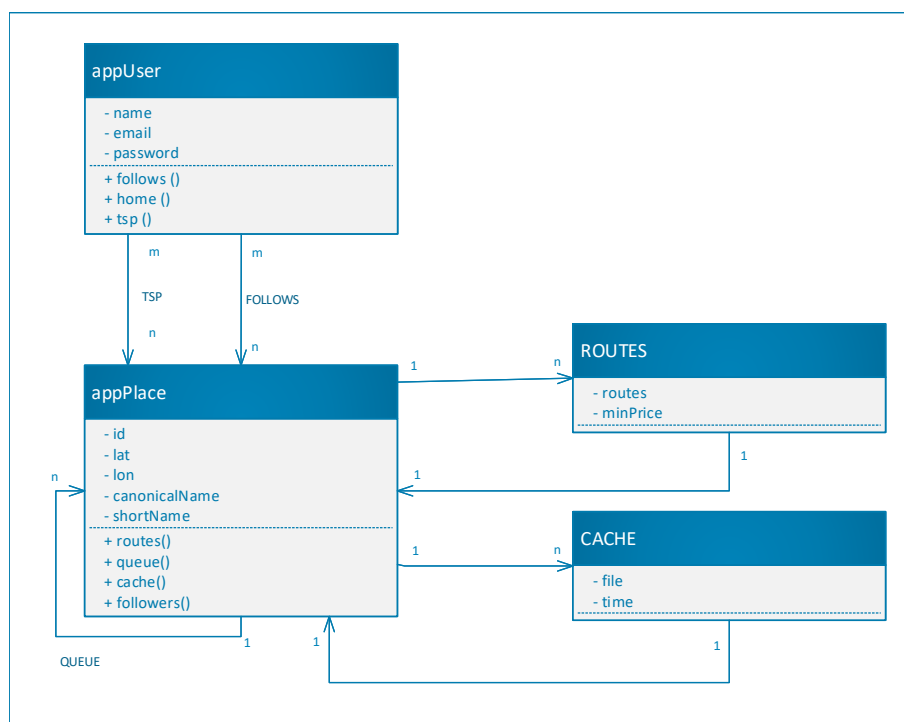
Obrázok 1: Usecase - Registrácia, autentifikácia a nastavenia



Obrázok 2: Usecase - dashboard



Obrázok 3: Usecase - tsp



Obrázok 4: Class diagram

5.2 Implementácia

5.2.1 Výber grafovej databázy

Vybrali sme si troch najpopulárnejších predstaviteľov grafových databáz podľa [1] rebríčka na DB-Engines.com. V nasledujúcej v skratke časti priblížime históriu každého GDBMS ich výhody, nevýhody pre naše použitie a proces výberu použitej databázy.

1. NEO4JPrvá verzia Neo4J bola vydaná v roku 2007 od vtedy sa stala dlhodobo najpoužívanejšou grafovou databázou. Je vyvíjaná Neo Technology, Inc. Neo4j je ponúkaná v dvoch variantách: Neo4j Community je open source (GPLv3) grafová databáza obsahujúca všetky základné funkcie (Ďalej budeme spomínať len túto verziu). A Neo4j Enterprise edícia, ktorá má rozšírené funkcie ako Shardovanie cache pamäte, rozšírené monitorovanie a zálohovanie za behu.
2. OrientDBOrientDb je vyvíjané od roku 2010 firmou OrientDB LTD. Databáza OrientDB rýchlo nabrala na popularite a v roku 2015 sa dostala na druhé miesto rebríčku db-engines. OrientDB sa rovnako ako Neo4j distribuuje v dvoch edíciach: Community - open source (Apache Licence 2.0) so základnými funkciami a Enterprise edíciou s podpodou migrácie a synchronizácie na Neo4J a pridanými

analytickými nástrojmi. + OrientDB podporuje okrem grafového modelu ukladania dát aj dokumentový a key/value model ukladania údajov.

3. Titan Titan bol od 2012 vyvíjaný skupinou thinkaurelius, no v roku 2017 bol odkúpený firmou Datstax a projekt Titan bol zastavený. Projekt je ďalej udržiavaný ako open source verzia pod menom JanusGraph. Titan je projekt určený na veľké distribuované enterprise riešenia, je nasadzovaný na cloudové platformy ako napr. Apache Hadoop, Apache Spark, ... podporuje rôzne distribuované úložné priestory ako napr. Apache HBase, Oracle BerkeleyDB, ... Ďalej podporuje rôzne vyhľadávacie enginy ako napr. Elasticsearch, Solr, ...
Toto riešenie je pre naše použitie nevhodné, lebo vyžaduje distribuovaný systém.

6 Implementácia

6.1 Použité technológie

6.1.1 Laravel framework

Laravel Framework je komplexný voľne šíriteľný framework. Tento framework je od roku 2015 najpopulárnejší PHP framework. Medzi jeho hlavné výhody patrí použitie relatívne novej verzie PHP 5.4, ktorá obsahuje technológie, ktoré v minulosti v PHP chýbali ako menové priestory a anonymné funkcie. Dalej obsahuje veľmi silný nástroj Eloquent ORM pre objektovo relačné mapovanie databázy, Blade šablónovací nástroj na rýchlu tvorbu dynamického obsahu.

Komplexita Frameworku Laravel je však aj jeho hlavnou nevýhodou, nieje vhodný na menšie projekty. V rýchlosti patrí medzi priemer PHP frameworkov.

6.1.2 NeoEloquent OGM

NeoEloquent OGM je voľne šíriteľná knižnica, ktorá umožňuje využívať grafovú databázu neo4j spolu s existujúcim dátovým modelom vo frameworku Laravel. Štruktúra NeoEloquent je modelovaná podľa existujúceho Eloquent Modelu a preto je jeho integrácia do ekosystému Laravel

6.1.3 Mapbox.js

Mapbox.js je komerčná knižnica na vytváranie projektov s interaktívnymi mapami. Je založená na voľne šíriteľnej knižnici Leaflet, rozširuje túto knižnicu o funkcie ako automatické zoskupovanie bodov do skupín a poskytuje bohatú a prehľadnú dokumentáciu. My používame verziu zdarma, ktorá je obmedzená na 50000 zobrazení mape na mesiac. Mapbox.js podporuje štandardný formát dát GeoJSON, tento formát umožňuje ukladať dáta o polohe, type bodu, rôznych atribútoch upresňujúcich vizuál zobrazovaného bodu. Tento formát ďalej umožňuje ukladať geometrické útvary a krivky.

6.1.4 Rome2Rio Api

Rome2Rio je portál ktorý zbiera údaje o cenách dopravy po celom svete a umožňuje vyhľadať cenu cesty medzi dvomi ľubovoľnými destináciami. Rome2Rio taktiež poskytuje niekoľko otvorených a platených API.

6.2 Vzhľad

Dôležitou požiadavkou kladenou na rozšírenie bolo príjemné používateľské rozhranie. Z tohto dôvodu malo rozšírenie obsahovať modifikovaných vlastností a tlačidlo pre prístup k nastaveniam rozšírenia v jednoduchšej a praktickej forme. Predpokladaný vzhľad je zo-

brazený na obrázku č. ??.

[?] Z tohto dôvodu malo rozšírenie obsahovať zoznam modifikovaných vlastností a tlačidlo pre prístup k nastaveniam rozšírenia v jednoduchšej a praktickej forme. Predpokladaný vzhľad je zobrazený na obrázku č. ??.

Záver

Cieľom práce bola analýza anonymizačných modulov, identifikačných prvkov prehliadača a vytvorenie anonymizačného modulu pre internetový prehliadač.

Analýzou najpoužívanějších modulov a vlastností prehliadača, ktoré slúžia na identifikáciu používateľa, sme zistili aktuálny stav a funkcionality rozšírení, ktorými je možné anonymizovať prístup na internet. Väčšina týchto rozšírení modifikuje len časť vlastností prehliadača, ktoré sú odosielané na server, alebo úplne blokuje ich odosielanie. Nami vytvorené rozšírenie dokáže modifikovať väčšinu identifikačných prvkov rozšírenia, pričom dodržiava súvislosti medzi vlastnosťami (používateľský agent odosielaný v hlavičke dopytu je totožný s používateľským agentom zisťovaním pomocou JavaScript príkazu, súvislosť medzi šírkou a dĺžkou rozšírenia obrazovky). Dokáže blokovat údaje, ktoré sú posielané v otvorenej podobe na server a obsahujú informácie o identifikačných údajoch prehliadača, ktoré sa nedajú na úrovni rozšírení modifikovať.

Testovanie rozšírenia nám overilo funkčnosť a správnosť implementácie. Rozšírenie dokáže buď vždy, alebo v časových intervaloch modifikovať väčšinu charakteristických prvkov prehliadača odosielaných na server, a tým zvyšuje anonymitu používateľa.

Zoznam použitej literatúry

- [1] AT, S. I. Db-engines ranking - popularity ranking of graph dbms, 2017.

Prílohy