

Tím č. 10

Kontakt na tím: gappers-tp-fiit@googlegroups.com

Barbora Brocková absolvovala bakalárske štúdium na FIIT STU v Bratislave. Venuje sa softvérovej analýze, má profesijné skúsenosti ako softvérová analytička – firma Jumpsoft a tiež ako projektová manažérka – Tatrabanka. Absolvovala Summer Startup School v roku 2013, podieľala sa na projekte testingo.sk. Má skúsenosti s programovaním v jazyku Java a Python, technológiami JavaScript a Bootstrap. V projektoch preferuje pozíciu analytičky a dokumentaristky.

Matej Čaja absolvoval bakalárske štúdium na FIIT STU v Bratislave. V rámci bakalárskeho projektu sa venoval téme Personalizované vyhľadávanie v digitálnej knižnici. Venuje sa tvorbe webových aplikácií prostredníctvom frameworku Ruby on Rails, profesijné skúsenosti má aj s jazykom Java. V projektoch sa podieľal na všetkých fázach vývoja softvéru. Vo voľnom čase si rád prehľbuje profesijné vedomosti.

Martin Číž absolvoval bakalárske štúdium na FIIT STU. Počas štúdia vyskúšal rôzne technológie, no orientoval sa najmä na jazyk Java. V roku 2013 absolvoval letnú programátorskú školu Summer Startup School kde sa naučil vytvárať webové aplikácie v jazyku Python, pracovať v tíme a rozšíril si znalosti o metódach vývoja softvéru. Venuje sa vytváraniu webových aplikácií v Ruby on Rails. Preferuje riešiť backend aplikácie.

Peter Kysel' absolvoval bakalárske štúdium na Fakulte riadenia a informatiky Žilinskej univerzity. V rámci bakalárskeho projektu vyvíjal desktopovú aplikáciu pre stavebnú spoločnosť. Pracuje pre IT Crowd. Preferuje programovanie v Jave, tvorbu HTML stránok a databázu Oracle. Má tiež skúsenosti s C++, PHP a jazykom Scala.

Adam Lieskovský absolvoval bakalárske štúdium na FIIT STU. V rámci bakalárskeho projektu sa venoval podpore tvorby dopytu v systéme Annota. Táto práca povzbudila jeho už i tak existujúci záujem o webové aplikácie, najmä ich backendovú časť. Pracuje s Ruby on Rails, ale má skúsenosti aj s Pythonom – stojí za projektom testingo.sk. Pozná Javu, HTML/CSS, Bootstrap, Coffeescript, jQuery, git a v poslednom období objavuje AngularJS. Má profesné skúsenosti s relačnými databázami (Oracle, MySQL, Postgres), ale rovnako aj NoSQL technológiami - Elasticsearch, Redis.

Daniel Uderman absolvoval bakalárske štúdium na FIIT STU. V rámci bakalárskeho projektu sa venoval vysokovýkonnému distribuovanému počítaniu s využitím algoritmov inšpirovaných hmyzom. Profesionálne sa venuje tvorbe ecommerce stránok v PHP. V priebehu štúdia získal aj skúsenosti s programovaním v C/C++, Jave a v jazyku JavaScript. Má skúsenosti tiež s databázovým systémom MySQL.

Michal Viskup absolvoval bakalárske štúdium na FIIT STU. V rámci bakalárskeho projektu sa zaoberal využitím grafickej procesorovej jednotky pre podporu výpočtu pri segmentácii obrazu. Zaujíma sa o spracovanie obrazu, počítačové videnie a strojové učenie. Pracuje pre spoločnosť vyvíjajúcu informačné systémy pre oblasť e-government, poštových a bankových služieb. Profesionálne programuje v Jave, má skúsenosti s technológiami spĺňajúcimi Java Enterprise Edition špecifikáciu. V rámci projektov tiež programoval v C/C++. Voľný čas venuje priateľke a návštevám posilňovne.

Téma č.15: Dav proti vizuálnemu smogu

Téma je pre nás atraktívna najmä v dvoch rovinách. Prvá rovina je spoločenská - téma poukazuje na pretrvávajúci problém v našej spoločnosti. Tento problém bol dlhodobo, s výnimkou zainteresovaných aktivistov, na periférii záujmu a všeobecne diskutovanou témou sa stal najmä vďaka médiám až teraz, tesne pred komunálnymi voľbami. Máme chuť túto vlnu záujmu využiť a vytvoriť produkt, ktorý umožní verejnosti aktívne sa zapojiť do zberu podnetov. Toto azda zvýši šance, že táto téma neupadne po chvíli do zabudnutia. Motiváciu pracovať v nás vzbudzujú celosvetovo úspešné projekty ako trashout.me [1], ktorý umožňuje prostredníctvom aplikácie pre 3 mobilné platformy nahlasovať svojim užívateľom nelegálne skládky odpadu a tiež úspešný projekt celoslovenského významu Odkaz pre starostu [2].

Tému tiež vnímame v technologickej rovine, kde predstavuje výzvu najmä z hľadiska spracovania dát nazbieraných používateľmi (zaznamenávanie lokality, zoskupovanie duplicitných záznamov). Ako výzvu tiež vnímame dosiahnutie čo najvyššej miery automatizácie tohto spracovania, napríklad prostredníctvom niektorých metód strojového učenia (v prípade dostatočne veľkého objemu zozbieraných dát).

Po odbornej stránke ponúkame ľudí so skúsenosťami s Ruby on Rails (Adam, Matej a Martin). Barbora má okrem programovania skúsenosti aj s analýzou a tvorbou dokumentácie na profesionálnej úrovni. Michal má znalosti z oblasti spracovania a analýzy obrazu a rozpoznávania objektov, ktoré by vedel zúročiť pri zoskupovaní duplicitných záznamov a triedení alebo označovaní zozbieraných záznamov. V práci získal tiež okrajové skúsenosti s klient-server komunikáciou s využitím REST.

V prvej fáze projektu vidíme ako kľúčové zozbieranie dát. Tieto dáta umožnia ladiť proces spracovania dát, najmä filtráciu relevantných dát a zoskupovanie duplícít. V rámci spracovania je tiež kľúčová schopnosť určiť, či sa jedná o legálne dielo (nelegálne nie je na základe toho, že to tvrdí ten, kto ju prostredníctvom aplikácie nahlásil...). Následne vidíme priestor pre vytvorenie rozhrania, ktoré umožní prezentovať spracované dáta. Tu sa ponúka možnosť naviazať zverejnenie dát na zverejnenie komunikácie so samosprávou a umožniť tak verejnosti sledovanie aktivity samosprávy po nahlásení diela. Zaujímavou môže byť aj možnosť umožniť na základe GPS polohy užívateľovi nájsť nahlásené diela.

[1] <http://www.trashout.me>

[2] <http://www.odkazprestarostu.sk>

Téma č.10: Sofistikované spracovanie dát a ich prezentácia

Internet je zdroj pre získanie širokého spektra údajov z odborných článkov, novinových príspevkov, informačných správ, osobných údajov o osobách alebo firmách (resp. organizáciách), obrázkov a fotografií, údaje o aktivite používateľov, ako je obľúbenosť stránky a údajov v nej, rôzne štatistické dáta a tak ďalej. V súčasnosti existuje niekoľko nástrojov, ktoré ich analyzujú, ale každý je špecializovaný len na konkrétny typ problému. Ďalším zdrojom dát na spracovanie sú výstupy z lokálnych databáz, softvéru (napr. ekonomický), ktoré by v rozumnom podaní vo forme prehľadných informácií umožnili bežným používateľom alebo aj malým podnikom profitovať zo získaných dát, priniesť pridanú hodnotu a zefektívniť ich prácu. Z toho dôvodu považujeme vytvorenie komplexného nástroja pre analýzu dát za jednu z perspektívnych tém, o ktorú bude záujem vo viacerých sektoroch spoločnosti.

Odraziac sa od predpokladu, že používateľovi umožníme načítanie dát z prehliadača alebo priamy import dát rôzneho typu a štruktúry, vedeli by sme si predstaviť niekoľko prístupov pre analýzu dát zo získaného datasetu. Nástroj by bol postavený na 3 základných typoch analýz, ktoré by uspokojili potreby používateľov z rôznych oblastí:

Prvým typom analýzy budú tradičné funkcie pre skúmanie textov dokumentov, ako sú analýza podobnosti dokumentov, zhlukovanie dokumentov do tematických množín podľa rôznych algoritmov, prehľadávanie nad dokumentmi podľa ich kľúčových slov spolu s používateľským hodnotením ich relevantnosti a odbornosti. (pod pojmom dokument môžeme označiť napr. jeden databázový záznam alebo jeden atribútov z neho). Dvaja členovia nášho tímu sa vo svojom bakalárskom projekte venovali určovaniu podobnosti dokumentov.

Druhým typom analýzy budú funkcie, ktoré spájajú širšie súvislosti medzi dátami vo vstupnom datasete. Patrí medzi ne prepájanie miest s náhľadmi v Google Maps (Street View), mien fyzických alebo právnických osôb s bližšími bibliografickými údajmi, súvislosti medzi dátami, ktoré ich spájajú cez osoby, miesta a čas. Na tieto features môžeme použiť verejne dostupné API týchto služieb. Nástroj sa bude učiť, ktoré údaje spolu súvisia a pri ich nasledujúcom identifikovaní ich odporučiť používateľovi s predpokladom vylepšovania sa nástroja s vzhľadom k objemu spracovaných dát.

Tretím typom analýzy budú vyhodnocovacie štatistiky nad tabuľkovými údajmi. Používateľom okrem štandardne známych číselných a reťazcových funkcií z tabuľkových editorov chceme ponúknuť previazanie s hore popísanými modulmi.

Náš tím má skúsenosti s vývojom webových aplikácií, pričom traja členovia priamo s Ruby on Rails. Všetci máme praktické skúsenosti s relačnými databázami, ale zopár členov aj s NoSQL riešeniami (napr. Elasticsearch), prípadne aj s BI. Motivuje nás aj množstvo príležitostí pre zapojenie ďalších moderných technológií (vizuálna prezentácia – interaktívne mapy, grafy).

Najväčšiu výzvu vidíme vo využití vhodných algoritmov a zapojenia strojového učenia, ktoré prinesú nástroju výhodu voči konkurencii, a to automatizované a inteligentné spracovanie dát. Veľký dôraz je nutné klásť aj na prívetivosť používateľského rozhrania, pretože potenciálnymi používateľmi môžu byť výskumníci z rôznych sfér alebo ľudia z praxe mimo IT sektora.

Téma č.8: Platforma pre monitorovanie chýb a iných udalostí v aplikácii

Možnosť monitorovania svojich nasadených aplikácií a upozornení na problémy vo forme chýb alebo bottleneckov je veľmi vítaná funkcionálna pre každého programátora, resp. správcu aplikácie. Po osobnej konzultácii sme boli usmernení v tom, v čom by mala aplikácia vynikať od už existujúcich riešení. Pri nich chýba identifikácia používateľov, ktorí s aplikáciou interagujú. Máme veľké množstvo údajov o používateľoch a ich správaní, dokonca v reálnom čase, ale nevieme priradiť jednotlivé akcie používateľom. Okrem možnosti definície vlastných udalostí vidíme túto možnosť ako veľmi plodnú pre následné využitie.

Vedeli by sme tak identifikovať aj vzory správania sa používateľov v našej aplikácii, čo by viedlo k novým možnostiam jej vylepšenia, prípadne zefektívnenia. Ďalej napr. osobnejší prístup k používateľom aplikácie, kde by sme vedeli pri vhodne zadefinovanej udalosti identifikovať problém, napr. aj neschopnosť/nepochopenie používateľa a následne ho kontaktovať a pomôcť mu alebo sa ospravedlniť za spôsobené problémy, ak chyba nastala na strane servera.

Jedno z prvotných riešení identifikácie používateľov v aplikáciach kde absentuje používateľský účet, je využitie cookies v prehliadači previazané na rýchle úložisko napr. Redis. Pokročilejším spôsobom môže byť identifikácia webového prehliadača, ktorá sa v práci Peter Esckersleyho ukázala ako relatívne funkčná [1].

Určite je cieľom nášho tímu ísť ďalej ako úroveň zachytávania chýb pre aplikácie na pôde fakulty. Veľmi radi by sme aplikáciu sprístupnili aj s benefitmi diskutovanými v predchádzajúcej časti, a to verejne a spravili ju ľahko použiteľnú pre bežných ITčkárov. Tým pádom chceme aplikáciu od začiatku navrhovať ako škálovateľnú, a poskytnúť verejne dostupné API, aby sa dali informácie získané z aplikácií integrovať a využiť aj individuálnymi spôsobmi pre jednotlivé aplikácie.

Všetci členovia nášho tímu majú skúsenosti s webovými MVC frameworkami, pričom Adam, Matej a Martin priamo s Ruby on Rails. Adam používa Newrelic aj Google Analytics, čo sú aplikácie z ktorých môžeme čerpať inšpiráciu.

[1] Peter Eckersley. 2010. How unique is your web browser?. In Proceedings of the 10th international conference on Privacy enhancing technologies (PETS'10), Mikhail J. Atallah and Nicholas J. Hopper (Eds.). Springer-Verlag, Berlin, Heidelberg, 1-18.

Príloha A

1. Dav proti Vizuálnemu Smogu
2. Sofistikované spracovanie dát a ich prezentácia
3. Platforma pre monitorovanie chýb a iných udalostí v aplikácii
4. Calendar People Matcher
5. Distribuovaný výpočtový systém – webové rozhranie
6. Inteligentný mobilný asistent
7. Big data kvality a ich využitie pri validácii a obohac.exist. dát
8. Testovanie použ. zážitku pomocou sledovania pohľadu
9. Vizualizácia informácií v obohatenej realite
10. Eyeblink - vyhod. frekvencie žmurkaní používateľa
11. Portál pre podporu rozvoja informatiky
12. Virtuálna FIIT
13. Hra pre mobilné zariadenia šitá na mieru osobnosti hráča