

Agregácia a odporúčanie miest a udalostí z Facebooku

Matej Dubeň*



Forget the boredom!

Let's Go Out!

Abstrakt

Cieľom tohto článku je priblížiť návrh aplikácie Let's Go Out! pre mobilné telefóny s operačným systémom Android, ktorá vznikla vrámci bakalárskej práce autora tohto článku. Zaobráva sa inteligentným a cieleným odporúčaním. Autor popisuje motivácie pre vznik aplikácie, samotný návrh a zaujímavé časti implementácie. Obsah článku je zameraný na odporúčacie systémy a podrobne vysvetluje ich funkcionality ako vo všeobecnej rovine, tak v konkrétnej rovine, ktorá spadá pod vývoj aplikácie Let's Go Out!. Vyvinutá aplikácia je voľne prístupná na stiahnutie. Avšak kvôli kontinuálnemu vývoju sa aplikácia dynamicky mení ako po grafickej stránke tak aj po funkcionálnej stránke, čo dočasne znemožňuje hĺbkové testovanie.

Kľúčové slová: Kolaboratívne odporúčanie — Obsahové odporúčanie — Agregácia udalostí — Agregácia miest — Android — Data mining — Kosínusová miera podobnosti

Priložené materiály: N/A

*xduben01@stud.fit.vutbr.cz, Faculty of Information Technology, Brno University of Technology

1. Úvod

Sociálna sieť Facebook je v dnešnej dobe veľmi populárna. Milióny ľudí ju využívajú ako zdroj informácií o miestach, prípadne udalostiach, v ich okolí. Avšak ľudia, ktorí stále rázne odmietajú sociálne siete, sa ku takýmto informáciám vôbec nedostanú, prípadne často až veľmi neskoro.

Popularita siete Facebook postupne narastá aj vďaka technologickému pokroku na poli mobilných zariadení. So zvýšujúcim sa počtom chytrých telefónov a tabletov, si už dnes veľa ľudí nedokáže život bez nich predstaviť. Mladšie generácie v mobilných zariadeniach vidia nekonečnú zábavu, zatiaľ čo staršie generácie ich využívajú skôr z pracovných dôvodov.

Podľa výskumu, ktorý urobila firma eMarketer¹,

bolo celosvetovo 1,75 miliardy používateľov chytrých telefónov. Z celkového počtu mobilných zariadení, z hľadiska operačného systému, má jednoznačne prvenstvo Android OS s podielom na trhu 76,6%, na konci roku 2014².

Na základe uvedených faktov vznikla aplikácia pre Android OS - *Let's Go Out!*, ktorá podporuje informovanosť, o lokálnych zaujímavých miestach a udalostiach, ľudí, ktorí sa sieti Facebook využívajú. Vzhľadom na fakt, že na sieti Facebook je obrovské množstvo údajov, ktoré často môžu užívateľovi nevyhovovať, je táto aplikácia postavená na **odporúčacom systéme**, ktorý zohľadňuje užívateľove záľuby a jeho zaznamenané správanie v minulosti.

Cieľom aplikácie je teda zapôsobiť na užívateľa

¹<http://goo.gl/EMKejm>

²<http://www.idc.com/prodserv/smartphone-os-market-share.jsp>

príjemným grafickým rozhraním a následne vyvolať vňom dojem, že aplikáciu *zaujímajú* užívateľove záujmy. Vyvolanie tohto pocitu je hlavným cieľom odporúčacieho systému. A nie hocjakého! Predpokladom je, aby systém dokázal spracovať ako údaje o jednotlivých položkách (udalosti a miesta), tak aj údaje o užívateľovi samotnom, a aby dokázal tieto údaje reflektovať pri výbere relevantných položiek. Toto všetko musí byť vykonávané v reálnom čase aby reakčná doba aplikácie bola čo najmenšia.

1.1 Konkurenčné aplikácia

Obchod s aplikáciami pre Android zariadenia (Google Play) obsahuje množstvo aplikácií, ktoré sa venujú problematike agregácie a odporúčania udalostí a miest. Na trhu je teda pomerne veľa konkurenčných aplikácií. Využívajú najmä údaje z Facebook profilu užívateľa, čo znamená, že užívateľ je nútenej sa do aplikácie prihlásiť pomocou svojho Facebook účtu. Tu ale nastáva problém v prípade, že užívateľ Facebook účet nemá. Na neho sa potom odporúčania nevzťahujú.

Tabuľka 1. Výhody a nevýhody existujúcich riešení oproti aplikácií *Let's Go Out!*

| All Events in City | |
|---|--|
| Výhody | Nevýhody |
| Vlastný webový portál | Necielene odporúčanie |
| Široká komunita | Agregácia podľa popularity |
| Interný zdroj udalostí | Neodporúča miesta |
| | Chýba vizualizácia na mape |
| Nearify | |
| Vlastný webový portál | Necielene odporúčanie Agregácia podľa popularity Chýba vizualizácia na mape Neodporúča miesta |
| Vamos | |
| Vlastný webový portál Viac externých zdrojov | Necielene odporúčanie Agregácia podľa popularity Neodporúča miesta Nepokrytie malých miest |
| Eventbride | |
| Vlastný webový portál Správa vstupeniek | Žiadne odporúčanie Neodporúča miesta Nevyužíva Facebook údaje Chýba vizualizácia na mape |

V tabuľke 1 je vidieť oblasti, v ktorých aplikácia *Let's Go Out!* vyniká oproti ostatným aplikáciám, rovnako ako aj oblasti, v ktorých oproti ostatným aplikáciám zaostáva. Vzhľadom na silu konkurencie musí teda byť riešenie správne navrhnuté a vysoko

optimalizované. Ako odporúčací systém bol teda vybratý koncept hybridného systému, ktorý zahŕňa ako odporúčanie na základe obsahu, tak aj kolaboratívne odporúčanie, ktoré berie do úvahy jednotlivé hodnotenia udalostí a miest a hľadá tak najbližšieho suseda ku aktuálnemu užívateľovi. Tento koncept je obohatený taktiež prvками strojového učenia a odporúčania so znalostným základom. Získaním užívateľových záujmov a sledovaním jeho správania si systém vybuduje bázu kľúčových slov a parametrov, ktoré potom ovplyvnia celkový výber odporúčených položiek. Keďže užívateľa najviac osloví vizuálna stránka aplikácie, je teda intuitívny grafický design nemenej dôležitý.

Do oblasti agregácie zaujíavých miest a udalostí teda prichádza aplikácia, ktorá dokáže vyhľadať a odporučiť nie len udalosti ale aj miesta, a to nie len novým užívateľom ale aj stálym užívateľom, ktorých záujmy aplikácia podrobne študuje a využíva pre osobnejšiu užívateľskú skúsenosť.

2. Úvod do odporúčacích systémov

Odporúčacie systémy vznikli za jediným účelom, a to pomôcť všetkým užívateľom nájsť vhodné a hlavne pre užívateľa relevantné informácie. Môžeme teda povedať, že odporúčacie systémy sú srdcom aplikácie, ktorá komunikuje priamo s užívateľom, spracováva spätnú vazbu, monitoruje užívateľove interakcie s aplikáciou a reflektuje tieto informácie do presnejších predpovedí.

Odporúčacie systémy môžeme rozdeliť, na základe zdroja informácií, podľa ktorého dochádza ku odporúčaniu, na 2 skupiny:

Obsahové odporúčacie systémy Dochádza ku hľadaniu najrelevantnejších informácií, na základe ich obsahu a kľúčových prvkov.

Kolaboratívne odporúčacie systémy Ku odporúčaniu využíva hodnotenia celej skupiny ľudí. Odporúčanie je presnejšie avšak nastáva problém, nazývaný "cold start" [1].

Pri **odporúčaní na základe obsahu** je predpokladom získať podrobne informácie o spracovávaných položkách. *Príklad:* Relevantné informácie pre knihy môžu byť *autor*, *žánor*, *kľúčové slová*, atď. Aplikácia *Let's Go Out!* využíva popis a názov miest a udalostí ako relevantný zdroj informácií.

Získavanie takýchto údajov môže byť ale veľmi pracné, a často aj problémové, pokiaľ dané informácie nie sú verejne prístupné. Systém taktiež musí tieto údaje zachovávať vo svojej databáze, čo pri veľkom počte položiek môže spôsobiť vysoké pamäťové nároky.

Na druhú stranu, pre správnu selekciu odporúčačných položiek nie je nutné zvažovať širokú škálu užívateľov a ich hodnotenia jednotlivých odporúčaných prvkov v systéme. Tieto systémy riešia otázku podobnosti prvkov, ktoré užívateľovi predložené neboli, s prvkami, ktoré užívateľ v minulosti hodnotil. Táto podobnosť je získavaná rôznymi prístupmi, medzi ktoré patrí napr. porovnávanie podobnosti atribútov (metadáta o prvku) alebo porovnávanie podobnosti obsahu textu na základe textovej klasifikácie a extrakcie kľúčových slov.

Narozdiel od obsahových odporúčacích systémov, ktoré pre potreby výpočtu odporúčania používajú podrobne údaje o prvkoch, sa **kolaboratívne systémy** spoliehajú na užívateľov v komuniti, ktorá aplikáciu používa. Základným princípom týchto systémov je nájsť užívateľa, ktorý je podobný v minulých hodnotniach s aktuálnym užívateľom aplikácie, pre ktorého sú aktuálne počítané položky ku odporúčeniu a cielene navrhnuté položky, ktoré aktuálny užívateľ ešte nevidel ("podobný" užívateľ ich v minulosti ohodnotil).

Pre potreby odporúčania je potrebná explicitná spätná väzba v podobe hodnotenia. Tento fakt môže spôsobovať problémy, pretože nie každý užívateľ je ochotný ohodnotiť svoju skúsenosť, čo znamená že nárast databázy hodnotení je veľmi pomalý. Taktiež tu čelíme problému "*cold start*", čo znamená že kolaboratívny odporúčací systém je závislý na databáze hodnotení, ktorá ale pri spustení systému neexistuje. Na internete je možné nájsť niekoľko dátových sád, ktoré môžu problém pomalého štartu zlepšiť. Tieto súbory údajov obsahujú základnú databázu položiek v systéme, sadu užívateľov a ich hodnotenie jednotlivých položiek (napr. filmová databáza *MovieLens*³).

Literatúra často uvádzá tretiu skupinu odporúčacích systémov, a to **hybridné systémy**. Hybridné odporúčacie systémy sú vytvorené kombináciou 2 predchádzajúcich skupín. Je tak docielené odstránenie nevýhod oboch skupín za cenu mierneho nárostu systémových nárokov. Odporúčania sú teda presnejšie ako pre nových užívateľov, tak aj pre stálu komunitu.

3. Návrh architektúry a teoretické riešenie problému

Vzhľadom na prácu s mobilnými zariadeniami bolo nutné zohľadniť pamäťové a výkonnostné kapacity cieľových zariadení a požiadavky na mobilnú sieť. Kvôli vyššiemu požadovanému výpočetnému výkonu pri cielenom doporučovaní bola vybraná architektúra zahŕňajúca *tenkého klienta a server*, s ktorým bude

³<http://grouplens.org/datasets/movielens/>

klient komunikovať na báze definovaných REST služieb. Server má teda špecifické aplikačné rozhranie, ktoré nemusí byť naviazané len na klienta v tomto projekte ale môže byť použité v ďalších aplikáciach.

3.1 Klientská časť - Android aplikácia

Klientská časť je písaná v jazyku Java za pomocí balíku vývojových nástrojov *Android SDK*. Ku Marcu 2015 bolo na trhu viac ako 92% zariadení, ktoré disponujú verziou Android OS 4.0 a vyššie⁴. Vzhľadom na tento štatistický fakt, je aplikácia vytvorená pre zariadenia s práve Android OS 4.0 a vyššie.

Android aplikácia slúži na vizualizáciu dát, ktoré sú spracované na serverovej časti. Tieto dátá prichádzajú vo formáte JSON, následne sú spracované a uložené do internej databázy na Android zariadení, pre rýchli prístup k už vyhľadaným údajom, v prípade nedostupnosti internetového pripojenia. Pre jednoduchosť a malé požiadavky bola vybraná technológia SQLite, na ukladanie dočasných dát v Android zariadení.

užívateľ poskytuje spätnú väzbu pomocou hodnotenia udalostí a miest, na škále 1 - 5. Toto hodnotenie je následne zaslanné na server, kde je pri kolaboratívnom filtrovaní využívané.

Ukážky výsledku návrhu sú zobrazené na obrázku 1. Avšak z dôvodu neustáleho vývoja aplikácie, sa reálne zobrazenie aplikácie môže od príkladných zobrazení lísiť.

3.2 Serverová časť - Odporúčací systém

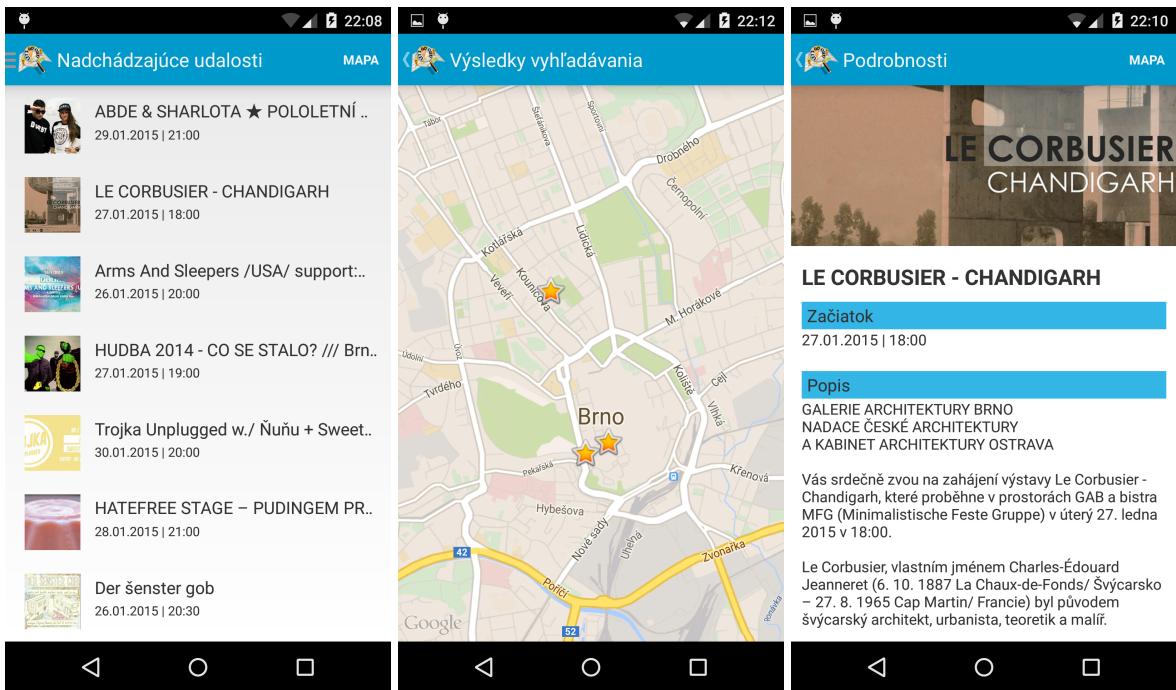
Serverová časť (a teda aj celý odporúčací systém) je napísaná v jazyku *Ruby on Rails*. Na základe Ruby on Rails architektúry (MVC⁵) je možné celú aplikáciu jednoducho rozšíriť o webový portál, ktorý bude splňať rovnakú funkcionality ako Android aplikácia a pritom môže využívať rovnakú business logiku ako využívajú aj REST služby (táto funkcionality je predmetom ďalšieho vývoja).

Ako už bolo v predchádzajúcich sekciách spomenuté (viz. podkapitola 1.1), po prieskume aplikácií s podobnou tématikou, bolo zistené, že existujúce aplikácie sa nevenujú cielenému odporúčaniu miest a udalostí užívateľovi. Cieľom tejto aplikácie je teda ponúknuť užívateľovi presnejší odhad miest záujmu a udalostí, ktoré by spadali do rozsahu jeho záujmov, než mu dávajú iné aplikácie.

Vzhľadom na to, že aplikačné rozhranie sociálnej siete Facebook ponúka detailné informácie ohľadom

⁴<https://developer.android.com/about/dashboards/index.html>

⁵Návrhový vzor užívateľského rozhrania, ktorý rozdeľuje funkcionality do 3 oddelených časťí (**Model**, **View**, **Controller**), ktoré medzi sebou komunikujú.



Obrázok 1. Názorné ukážky výsledku návrhu vzhľadu implementovaného v aplikácii.

miest a udalostí, sú tieto údaje spracovávané pre vytvorenie sady kľúčových slov a položky sú teda zaradené a ľahko dohľadateľné. Na princípe spracovania existujúcich detailných informácií je založená **obsahová časť odporúčania**. Pre získavanie kľúčových slov z popisu miest a udalostí sa používa kódovací formát **TF-IDF** [1] (*term frequency - inverse document frequency*). Text, kódovaný formátom TF-IDF, je reprezentovaný ako vektor jednotlivých váh pre vybrané kľúčové slová v texte. Na výpočet váhy kľúčového slova sa používa nasledujúci vzťah:

$$TF-IDF(k, d) = TF(k, d) * IDF(k) \quad (1)$$

Koeficient **TF** predstavuje frekvenciu kľúčového slova v spracovávanom teste. Avšak môže sa stať, že položky s dlhším spracovávaným textom budú mať vyššiu prioritu ako tie s kratším textom. Je preto nutné počítanie koeficientu TF normalizovať (viz vzťah 2).

$$TF(k, d) = \frac{frekv(k, d)}{maxOthers(k, d)} \quad (2)$$

Normalizácia v prípade použitého vzťahu spočíva vo využití maximálneho počtu výskytov iného, než spracovávaného, kľúčového slova. [2]

V spracovávaných informáciách sa často vyskytujú slová, ktoré majú vysoký počet výskytov, avšak nedajú sa radi medzi kľúčové slová (napr. spojky, predložky, atď.). Koeficient **IDF** slúži na eliminovanie vysokej váhy slov tohto typu. Počíta sa nasledovne:

$$IDF(k) = \log \frac{N}{n(k)} \quad (3)$$

Vypočítaný vektor TF-IDF vám je následne používaný vo vzťahu pre mieru podobnosti 2 záznamov (viz vzťah 4). V tomto prípade dochádza ku porovnávaniu textu popisu udalostí a miest, ktoré užívateľ už navštívil (zhliadol podrobnosť miesta/udalosti v aplikácii) a udalostí a miest, ktoré by mu mohli byť odporučené. Miera podobnosti následne určí relevantnosť danej položky voči aktuálnemu užívateľovi.

Druhou časťou navrhnutého odporúčacieho systému je **kolaboratívna časť**. Ako už bolo písané v kapitole 2, tento systém je citlivý na údaje o hodnotení od ostatných užívateľov. Presnosť odporúčania sa zvyšuje dlhším aktívnym používaním systému (predovšetkým hodnotenie položiek). Pri získavaní odporúčaných položiek je potrebné získať užívateľov, ktorí majú hodnotenie rovnakých udalostí a miest najbližšie ku aktuálnemu užívateľovi. Ako všeobecný štandard získavania miery podobnosti dvoch položiek v systéme sa používa **kosínusová miera podobnosti** (angl. Cosine Similarity Measure). Podobnosť sa počíta na základe uhlu medzi dvomi n -rozmernými vektormi (viz. 4).

$$sim(\vec{x}_1, \vec{x}_2) = \frac{\vec{x}_1 \cdot \vec{x}_2}{|\vec{x}_1| * |\vec{x}_2|} \quad (4)$$

Vektory x_1 a x_2 predstavujú vektory hodnotení 2 užívateľov, medzi ktorými hľadáme podobnosť. Rovnica nadobúda hodnoty v rozmedzí 0 a 1. Čím bližšie je hodnota rovná 1, tým podobnejšie sú 2 porovávané objekty.

4. Realizácia získavania dát a hybridného riešenia

Pre správnu funkčnosť serverovej strany aplikácie sú potrebné 2 kľúčové časti. *odporúčací systém a proces získavania údajov.*

4.1 Získavanie dát

Sociálna sieť Facebook zverejnila aplikačné rozhranie, pomenované **Graph API**⁶, pomocou ktorého je možné prehľadávať a získavať údaje o stránkach (reprezentujúcich miesta), udalostiach, atď. Ako už napovedá názov, Graph API reprezentuje dátá v podobe sociálneho grafu, kde jednotlivé záznamy sú rozdelené na uzly a hrany medzi nimi.

Odpoveďou na požiadavku, pomocou Graph API, je objekt, príp. pole objektov, vo formáte JSON, ktoré obsahuje všetky detailné údaje, ku ktorým je voľný prístup a o ktoré bolo zažiadane. Tieto údaje sú následne roztriedené a uložené do databázy, kde s údajmi pracuje už odporúčacia časť serverovej časti.

Pre prístup ku Graph API verejným údajom nie je užívateľ povinný prihlásiť sa, čo rozširuje možnú klientskú bázu o užívateľov bez prístupu ku sieti Facebook.

4.2 Hybridné riešenie

Za pomoci vzťahov, ktoré boli vysvetlené v podkapitole 3.2, bol vyvinutý ako obsahový odporúčací systém, tak aj kolaboratívny systém. Tieto 2 systémy spracovávajú extrahované údaje zo siete Facebook nezávisle na sebe. Ich výstup je definovaný ako súbor identifikátorov miest a udalostí, ktoré spadajú do rozmedzia záujmov užívateľa. Výstupy sú spojené do jedného celku, pričom duplicity sú odstránené, a takto vytvorený tok dát je zaslaný na užívateľove zariadenie, prostredníctvom REST⁷ služieb, kde sú tieto údaje patrične vizualizované.

Vzhľadom na spojenie výsledkov ako obsahového tak kolaboratívneho odporúčacieho systému, aplikácia prispôsobuje svoje správanie novým i zabehnutým užívateľom.

5. Záver

Cieľom tohto článku bolo priblížiť návrh aplikácie *Let's Go Out*, ktorej hlavným cieľom je priblížiť sa čo najviac k užívateľovi, zaznamenávať fakty z používania aplikácie a reflektovať ich do cieleného odporúčania,

⁶<https://developers.facebook.com/docs/graph-api>

⁷REST (Representational State Transfer) služby slúžia ku komunikácii medzi klientom a serverom na báze HTTP správ.

ktoré existujúcim aplikáciám chýba. Vypočítané odporúčania následne vizualizovať v užívateľsky príjemnej forme, aby čas k dosiahnutiu informácií bol čo najmenší.

Aplikácia je stále vo fáze kontinuálneho vývoja a neustále prebiehajú zlepšovania, ako grafického užívateľského rozhrania, tak aj implementácia a optimizácie na strane serveru. Napriek tomu je však aplikácia zverejnená a voľne dostupná na internetovom obchode s aplikáciami pre Android zariadenia, Google Play⁸. Avšak odporúčací systém neboli ešte podrobne testovaný, práve kvôli neukončenému vývoju niektorých základných častí.

Ako bude aplikácia súbojiť, s ostatnými konkurenčnými riešeniami, sa ukáže v blízkej budúcnosti, kedy je taktiež naplánovaný vývoj webového portálu, ktorý bude slúžiť pre rovnaký účel, ako samotná aplikácia.

Podakowanie

Dakujem svojmu vedúcemu bakalárskej práce, Ing. Igorovi Szőkemu, Ph.D., za cenné rady a pripomienky.

Literatúra

- [1] Alexander Felfernig Gerhard Friedrich Dietmar Jannach, Markus Zanker. *Recommender Systems, An Introduction.* Cambridge University Press, 2011. ISBN: 978-0-521-49336-9.
- [2] Soumen Chakrabarti. *Mining the Web: Discovering Knowledge from Hypertext Data.* Morgan-Kauffman, 2002. ISBN: 1-55860-754-4.

⁸Aplikáciu nájdete na adrese <https://play.google.com/store/apps/details?id=cz.vutbr.fit.letsgoout>