

Adaptívna podpora spolupráce v komunitnom systéme otázok a odpovedí

Marek Grznár
Vedúci: Ivan Srba

CQA

- Comunity Question Answering
- Nieкто sa spýta otázku na svoj problém a komunita mu odpovedá.
- Stack Overflow, Quora, Yahoo! Answers atd'...
- Cieľ CQA → poskytnúť odpoveď na otázku v čo najkratšom čase

Question routing

- Hľadanie vhodných používateľov, ktorí by na otázku vedeli odpovedať.
- Súčasné prístupy → využívajú iba dáta zo systému (otázky, odpovede...) = QA dáta

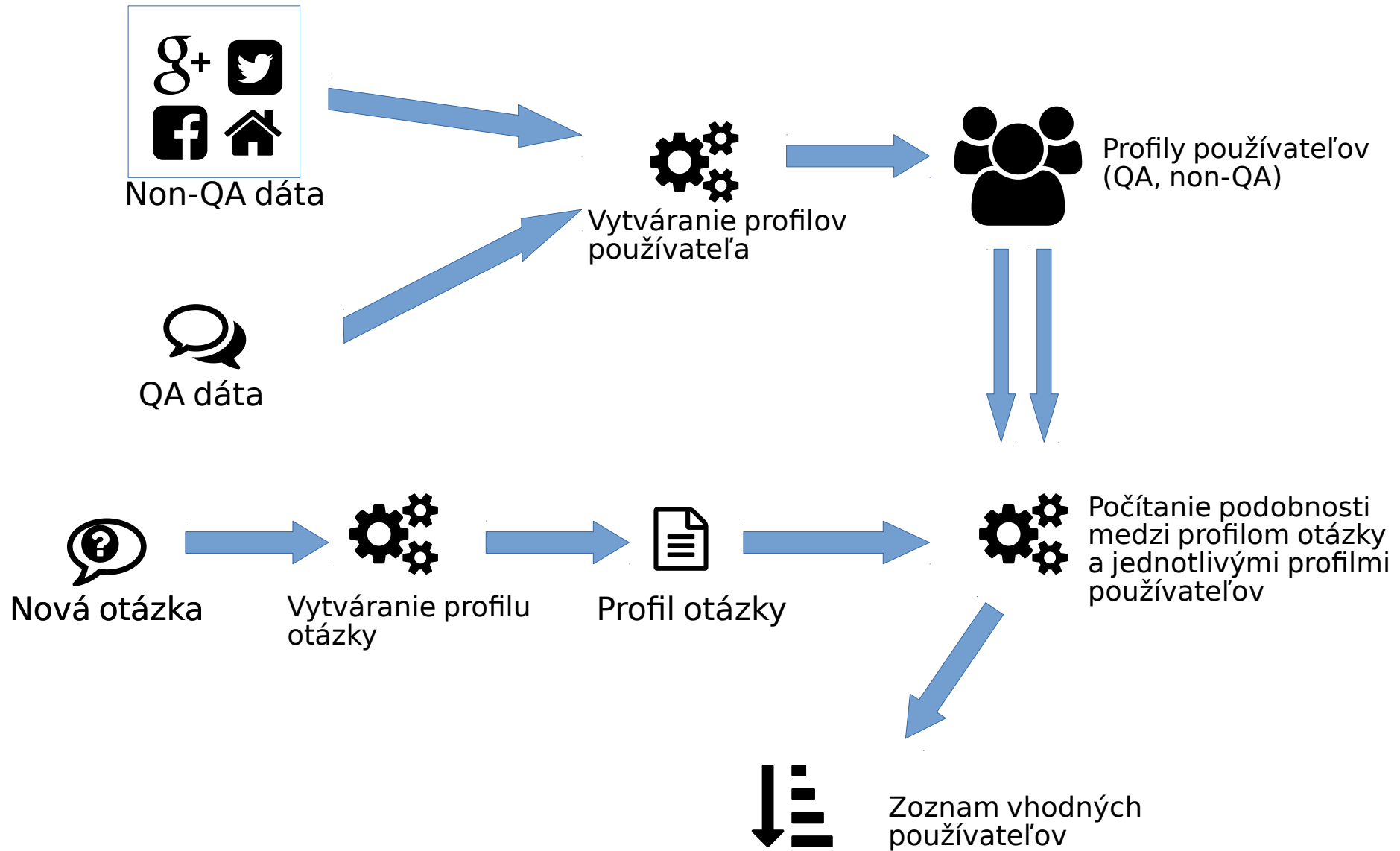
Problém

- Slabá história používateľovej aktivity
- Lurkers – používatelia, ktorí sa málo zapájajú do riešenia otázok
- Cold start problem

Hypotéza 1

- Využitie non-QA dát pomôže v systéme odhaliť používateľov, ktorí sú ochotní poskytnúť odpoveď na otázku.
- Non-QA dáta → dáta, ktoré používateľ o sebe voľne šíri (blogy, twitter, FB priatelia...)
- Offline overenie na dataseze zo Stack Overflow

Návrh metódy



Profil používateľa

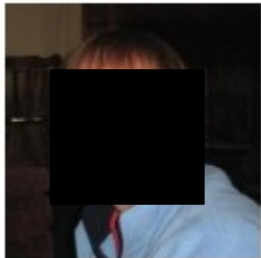
- Profil používateľa z QA dát
 - Využitie existujúceho prístupu na základe LDA
 - Profil používateľa → zachytáva expertízu (interest) používateľa o nejakú tému
- Profil používateľa z non-QA dát

Profil používatele'a z non-QA dát

- Home page, About me, Twitter → ako samostatné dokumenty

Bryan Denny [less info](#)

 meta user |  network profile



13,474

reputation

● 20 ● 79 ● 112

<i>bio</i>	website	bryandenny.com
	location	Winston-Salem, NC
	age	27
<i>visits</i>	member for	6 years, 3 months
	seen	6 hours ago
<i>stats</i>	profile views	1,516

Currently web (C#/ASP.NET) programmer for a printing business.

Former computer science college student with a strong background in Java and C#.

I'm a **moderator** on the [Android StackExchange](#). I own a G1 running CM5 and a Motorola Droid running CM6. I also own a rooted Nook Color and a wifi Motorola Xoom. I currently use a rooted Galaxy Nexus as my main device.

You can also find me on [Twitter](#).

I'm also a new Android developer. Check out my tipping calculator app called "Tippy Tipper" or my car dock app called "Car Dashboard" [here](#). The source code for "Tippy Tipper" can be found [here](#).

Profil používateľa z non-QA dát

- Využitie LDA pre každý dokument
- Výsledný profil používateľa ako lineárna kombinácia tém z jednotlivých dokumentov
 - Možnosť experimentovať s váhami jednotlivých dokumentov

Profil otázky

- Pravděpodobnost, že otázka patří do témy
 - pravděpodobnosti jednotlivých slov v slovníku
 - Slovník = množina jedinečných slov vyskytujících v UP

Párovanie používateľa (UP) a otázky (QP)

- Porovnanie na základe kosínusovej podobnosti pravdepodobnosti výskytu jednotlivých slov v otázke/téme.
- Osobitné počítanie podobnosti medzi
 - QP a $UP_{\text{non-QA}}$
 - QP a UP_{QA}
- Výsledné párovanie
 - Lineárna kombinácia jednotlivých podobností
 - Hľadanie vhodného pomeru non-QA a QA profilmi

Experiment a vyhodnotenie

- Offline experiment
- Dataset Stack Overflow
- Porovnanie voči zlatému štandardu
 - Zlatý štandard: používatelia, ktorí skutočne odpovedali na otázku
 - Problém → neúplný dataset, chýbajúce záznamy o pozretiach otázky, hlasoch...

Hypotéza 2

- Použitie non-QA pomôže zrýchliť odpovedanie na otázky
- Online experiment
 - Askalot, implementované rozšírenie pre zber všetkých potrebných dát
- Využitie priateľstva na Facebook ako zdroj non-QA dát

Zhrnutie

- Smerovanie otázok s využitím non-QA dát
- Budovanie expertízi používateľa z voľne dostupných dát
- Vytváranie modelov (používateľa, otázky) pomocou LDA
- Porovnanie podobnosti pomocou kosínusovej podobnosti
- Offline overenie – Stack Overflow
- Online overenie – Askalot
 - Implimentovaný zber dát