

Reputácia používateľov v systémoch pre odpovedanie na otázky v komunitách

Adrián Huňa

Vedúci práce: Ing. Ivan Srba

17. jún 2015

Systemy pre odpovedanie na otázky v komunitách (CQA systémy)

- ⦿ Angl. **C**ommunity **Q**uestion **A**nswering systems
- ⦿ StackOverflow, Yahoo! Answers, Quora, Askalot, ...
- ⦿ Cieľ: Vytváranie centier vedomostí
 - > Ako a kto ich vytvára?



- > Ako vieme, ktorí používatelia vytvorili najviac vedomostí?

Reputácia v CQA systémoch

- Odráža prínos používateľa pre komunitu
- Kombinácia **expertízy** a **aktivity** používateľa
- Bodový systém

asked Nov 7 at 1:26



4,367 ● 2 ● 22 ● 47

Identifikovanie expertov

⊙ Grafové prístupy

> Grafy

- Pýtajúci sa – odpovedajúci
- Pýtajúci sa – najlepší odpovedajúci
- Najlepší odpovedajúci – ostatní odpovedajúci

> Algoritmy

- Degree
- HITS
- PageRank

⊙ Prístupy založené na vlastnostiach

> Vlastnosti používateľov

- Počet otázok, počet odpovedí, ...

> Vlastnosti otázok

- Hodnotenie, počet odpovedí, ...

> Vlastnosti odpovedí

- Hodnotenie, status najlepšej odpovede

Problémy identifikovania expertov

- ⊙ Rôzna náročnosť a kvalita otázok a odpovedí
- ⊙ Ignorovanie rozdielov medzi témami CQA systémov
- ⊙ Existujúce prístupy dávajú väčšiu váhu aktivite používateľov
 - > My však potrebujeme poznať expertov => dať im väčšie práva, zvýrazňovať ich odpovede, smerovanie náročných otázok

Vlastnosti navrhnutej metódy výpočtu reputácie

- ⦿ Prístup založený na vlastnostiach
- ⦿ Zameranie na expertízu používateľov
- ⦿ Zohľadnenie rôznej kvality a náročnosti otázok a odpovedí
- ⦿ Zohľadnenie rozdielov medzi témami CQA systémov
- ⦿ Uvažovanie odpovedania aj pýtania sa

Hypotéza

- *Ak pri identifikovaní expertov vezmeme do úvahy náročnosť otázok a užitočnosť otázok a odpovedí, budeme schopní expertov identifikovať s väčšou úspešnosťou než doterajšie jednoduché vlastnostné prístupy.*

Výpočet reputácie

- ⊙ Súčet čiastkových reputácií za pýtanie sa (R_q) a odpovedanie (R_a)

- > $R(u) = \sum R_q(q_u) + \sum R_a(a_u, q)$

- ⊙ Alternatívny tvar úplne eliminujúci aktivitu

- > $R(u) = \frac{\sum R_q(q_u) + \sum R_a(a_u, q)}{|q_u| + |a_u|}$

Reputácia sa otázky

$$R_q(q_u) = D_q(q_u) + QU(q_u)$$

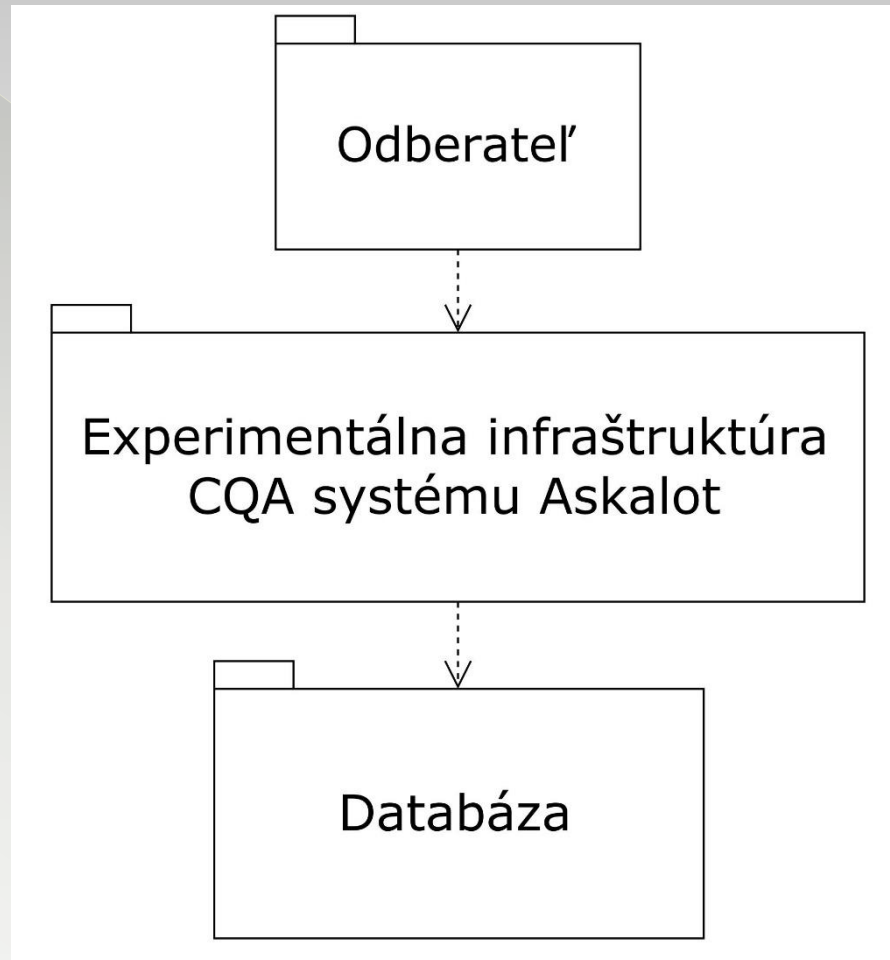
- ◎ Náročnosť otázky (D_q)
 - > Čas do pridania prvej odpovede
 - > Čas do akceptovania odpovede za najlepšiu
 - > Priemerný čas odpovedí
- ◎ Užitočnosť otázky (QU)
 - > Hodnotenie otázky hlasovaním
- ◎ Normalizovanie a logaritmovanie hodnôt
 - > Rozdielna populácia medzi témami a problém dlhého chvosta rozloženia hodnôt

Reputácia za odpovede

$$R_a(a_u, q) = D_q(q) + AU(a_u, q)$$

- ⊙ Náročnosť otázky
- ⊙ Užitočnosť odpovede
 - > Hodnotenie odpovede hlasovaním
 - > Status najlepšej odpovede, ak sú všetky odpovede s nulovým hodnotením
- ⊙ Normalizovanie a logaritmovanie hodnôt

Implementácia



Overenie

- ⦿ Porovnanie zoradenia odpovedajúcich používateľov pre každú otázku
 - > vypočítaná reputácia vs. hodnotenie odpovede
- ⦿ Programmers.stackexchange.com
- ⦿ 33 000 otázok
 - > Vyhodnotenú na 20 000 otázkach

⊙ Metriky:

- > P@n (P@1, P@2)
- > MRR
- > nDCG

⊙ Metódy

- > StackOverflow reputácia
- > Pomer najlepších odpovedí
- > Z-score
- > Počet odpovedí

	P@1 (%)	P@2 (%)	MRR (%)	nDCG (%)	# Otázok
Naša metóda (<i>súčet</i>)	40,093	38,074	65,538	83,162	20552
Naša metóda (<i>priemer</i>)	43,971	41,154	66,278	84,511	20552
StackOverflow reputácia	40,699	38,474	63,943	83,370	20558
Pomer najlepších odpovedí	41,881	40,078	64,585	83,728	20324
Z-score	38,388	37,022	62,322	82,534	20558
Počet odpovedí	38,570	37,308	62,481	82,647	20324

Aplikácia do Askalotu

- ◉ Výpočet reputácie
- ◉ Vizualizácia



Zhodnotenie

- ⦿ Výpočet reputácie používateľov zameraný na expertízu
- ⦿ Objektívne zachytenie množstva expertízy za jednotlivé otázky a odpovede
- ⦿ Eliminácia rozdielov tém v CQA systémoch
- ⦿ Naša metóda prekonala porovnávané metódy
- ⦿ Priestor na zlepšenie v objektívnosti výpočtu a určení váh jednotlivých zložiek reputácie