

Predpovedanie kľúčových slov do vyhľadávača na základe potreby používateľa odvodenej zo sledovania pohľadu

Autor: Patrik Januška

Vedúci práce: Ing. Eduard Kuric

Cieľ práce a hypotéza

Cieľ

Navrhnuť metódu na rozšírenie dopytu vyhľadávania s využitím sledovania (zaznamenávania) pohľadu, pričom výsledky vyhľadávania budú pre používateľa relevantnejšie s použitím rozšíreného dopytu v porovnaní s pôvodným dopytom.

Hypotéza

Navrhnutá metóda odporučí rozšírenie dopytu o kľúčové slová tak, aby výsledky vyhľadávania boli relevantnejšie pre používateľa, v porovnaní, keď nepoužije rozšírený dopyt.

Vyhľadávanie na Webe

Typy dopytov

- *Navigačné*
- *Transakčné*
- *Informačné*

Záujem používateľa

- Dlhodobý záujem používateľa (long-term interest)
 - *história navštívených stránok používateľom, zadávané dopyty, atď. (P.A. Chirita 2007)*
- Krátkodobý záujem používateľa (short-term interest)
 - *aktuálne navštívené stránky, hľadané informácie (D. Kelly 2004)*

Navrhnuté metódy a spôsob ich riešenia

- Východisková metóda
- Metóda filtrovania pohľadu
- Metóda filtrovania dĺžky pohľadu
- Metóda filtrovania dĺžky pohľadu – Rozšírená
- Metóda filtrovania dopytu

Východisková metoda

- Aplikujeme TF/IDF na všechny zachytené slova
- Východiskový dokument = obsah všech odsekov web stránky
- Klíčové slova

Metóda filtrovania pohľadu

- Aplikujeme TF/IDF na všetky zachytené slová
- Východiskový dokument = obsah prečítaných odsekov web stránky
- Kľúčové slová

Metóda filtrovania dĺžky pohľadu

- Prečítané odseky
- Dĺžka prečítaného odseku (kľúčových 230 znakov)
- Aplikujeme TF/IDF len na zachytené slová z prečítaných odsekov
- Východiskový dokument = obsah prečítaných odsekov web stránky
- Kľúčové slová

$$interest(t) = \frac{LA(t)}{LA(t) + SA(t)} * TF/IDF(t)$$

Metóda filtrovania dĺžky pohľadu - Rozšírená

- Rovnaké pravidlá ako v predchádzajúcej metóde
- Veľkosť zreničky
- Počet fixácií pre dané slovo

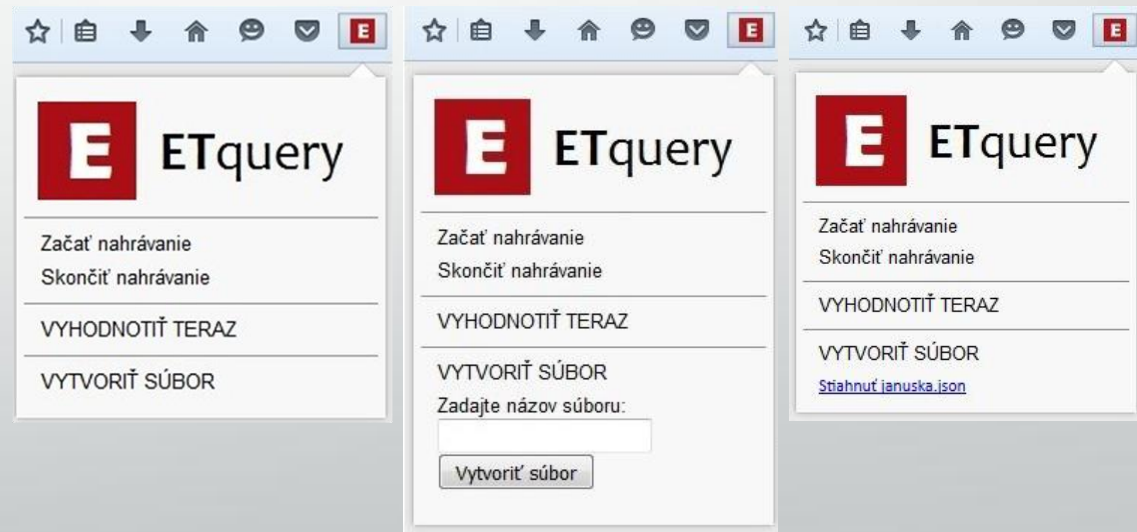
$$interest(t) = \frac{LA(t)}{LA(t) + SA(t)} * TF/IDF(t) * averagePS * fixationCount$$

Metóda filtrovania dopytu

- Aplikujeme TF/IDF na všetky slová zo všetkých odsekov
- Zhlukovanie slov do skupín
- Porovnanie dopytu so zhlukmi slov

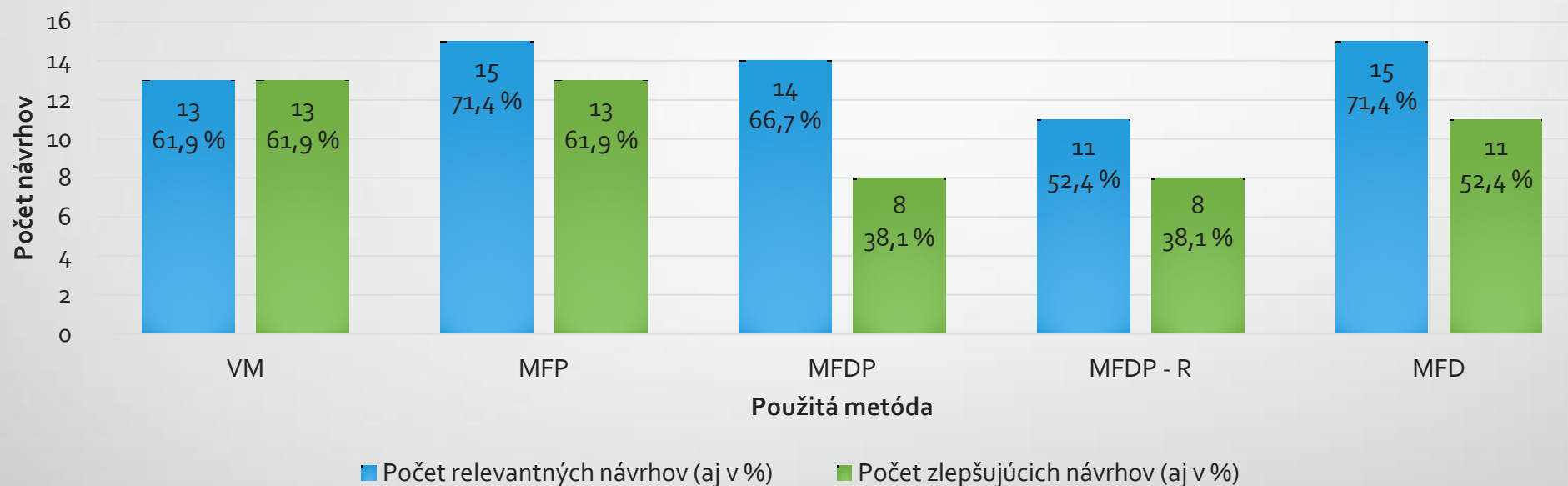
Experiment

- 7 účastníkov
- Článok (1800 slov) - atrakcie, dovolenkové destinácie, šport a kultúra
- Abstrakty k témam
- Označenie relevantných výsledkov
- Klient – server aplikácia
- Mozilla Firefox Rozšírenie



Vyhodnotenie

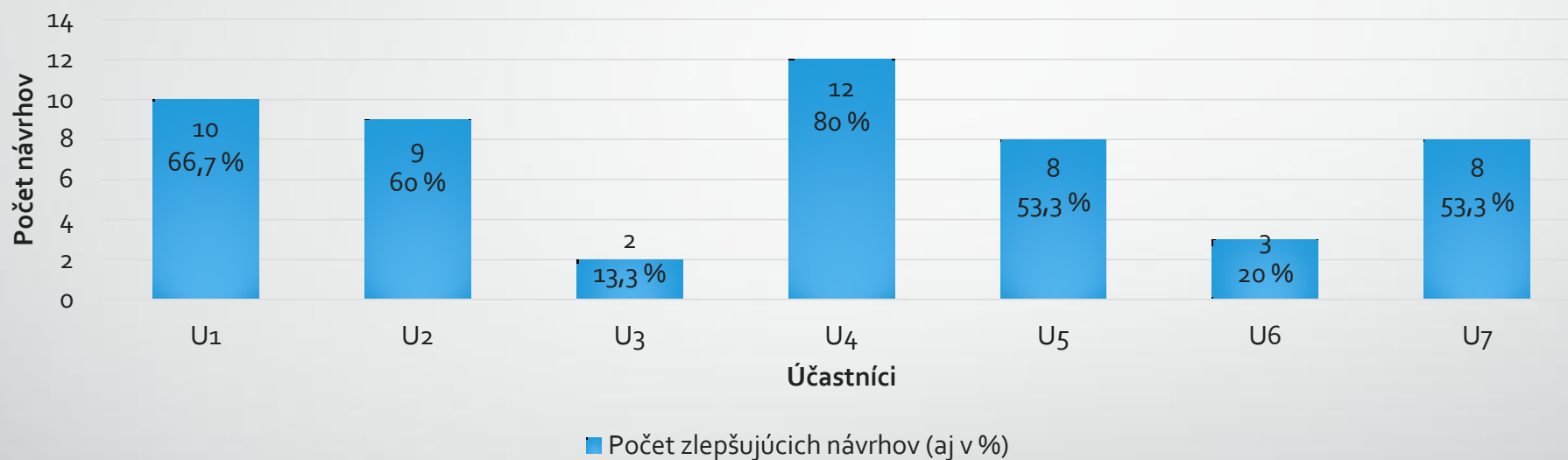
Prehľad počtu návrhov metód



- S našimi metódami dokážeme s priemernou presnosťou 64,8 % určiť vyhľadávací zámer.
- S našimi metódami dokážeme v priemere poskytnúť 50,8 % zlepšujúcich návrhov.

Vyhodnotenie

Prehľad počtu zlepšujúcich návrhov jednotlivých účastníkov



- V priemere 49,5 % návrhov pri každom účastníkovi prinieslo zlepšenie.

Zhrnutie

- Metóda na odporučenie rozšírenia dopytu o kľúčové slová tak, aby výsledky vyhľadávania boli relevantnejšie pre používateľa
- Informačné dopyty
- Krátkodobý záujem
- Zaznamenávanie a analyzovanie čítania textu používateľom
- Odvodenie záujmu používateľa z čítania