

Activity-based Search Session Segmentation

Samo Molnár

2014/04/14

Søke

- ▶ Search Logging Platform
- ▶ Google suggestions
- ▶ Bing results
- ▶ *Try it!*

<http://j.mp/soke-search>

Identifikácia vyhľadávacích epizód

- 1) Nájdí dopyty, ktoré zdieľajú aspoň jeden výsledok s aktuálnym dopytom.
- 2) Vypočítaj novú relevanciu pre zdieľaný výsledok na základe aktivity používateľa pre daný dopyt.
- 3) Podobnosť dopytu je suma relevancií zdieľaných výsledkov.
- 4) Vyber najrelevantnejší dopyt.

Approach

- ▶ Query to Query comparison
- ▶ Lexical similarity (Levenshtein, Cosine similarity)
- ▶ Shared results
- ▶ dwell time to increase relevance

$$s(Q_1, Q_2) = \log(|R_s|) + \sum_{(\forall R \in R_c)} \frac{\log(t(R)+1)}{\log(pos(R)+1)}$$

Approach

- ▶ Global/per user results occurrence not considered
- ▶ Negative feedback not considered as well

Evaluation

- ▶ SVM by query pairs
- ▶ TREC {2011, 2012, 2013}
- ▶ Sessions with at least one clicked result (157 s. \w 634 q. \w > 113k ex.)
- ▶ only lexical features: 76%
- ▶ lexical & shared results features: 78%

Søke statistics

- ▶ Queries with at least one click: 589 / 782
- ▶ Median of dwell time: 2m13s
- ▶ Shared results: 75 sessions
- ▶ Clicked positions:
 - ▶ 1th — 420
 - ▶ 2nd — 193
 - ▶ 3rd — 111

<http://j.mp/soke-search>

Plan

- ▶ Søke evaluation this week
- ▶ Other results features from Søke
- ▶ Yandex Switch Detection dataset
- ▶ Writing and finishing up...

<http://j.mp/soke-search>