Exploratory Search Missions for TREC Topics

Martin Potthast  Matthias Hagen  Michael Völkske  Benno Stein
Bauhaus-Universität Weimar
99421 Weimar, Germany

We report on the construction of a new text reuse corpus comprising writing interactions and exploratory search missions.

- 150 essays (based on TREC Web Track topics 2009-2011)
- 12 professional writers hired on a crowdsourcing platform
- Long essay writing task, researching sources using a custom ClueWeb09 search engine
- Writing and search engine interactions recorded in high detail

Data Collection

Search mission data will be made available as the Webis-Query-Log-12 (http://www.webis.de/research/corpora)

Main Findings

Spectrum of search behavior

- Percentage of queries submitted over time for all 150 search missions
- Ranges from majority of queries issued at the start of the task (A1) to most queries towards the end (J15)
- In between, sets of queries submitted in bursts (e.g. F9) or linear increase (A10)

Correlation of searching and writing

- Evidence of distinct text reuse strategies (build-up and boil-down)
- Only the former clearly reflected in the query log

First Conclusions

- Query frequency by itself poor predictor of task completion
- Heavy reliance on search engine indicates need to better support exploratory tasks