Head First: Living Labs for Ad-hoc Search Evaluation

Krisztian Balog  
University of Stavanger  

Liadh Kelly  
Trinity College Dublin  

Anne Schuth  
University of Amsterdam

Overall goal

• Make information retrieval evaluation more realistic
  Evaluate retrieval methods in a live setting with real users in their natural task environments

Our focus

• Medium-sized organizations with fair amount of search volume
  Typically lack their own R&D department, but would gain much from improved approaches

Key idea

• Focus on frequent (head) queries
  • Enough traffic on them (both real-time and historical)
  • Ranked result lists can be generated offline
  • An API orchestrates all data exchange between live sites and experimental systems

Methodology

• Queries, candidate documents, historical search and click data made available
• Rankings are generated for each query and uploaded through an API
• When any of the test queries is fired, the live site request rankings from the API and interleaves them with that of the production system
• Participants get detailed feedback on user interactions (clicks)
• Ultimate measure is the number of “wins” against the production system

Benchmarking campaign

• Three ad-hoc search tasks as use-cases
  • Product search (on an e-commerce site)
  • Local domain search (on a university’s website)
  • Web search (through a large commercial search engine)
• To be extended with additional tasks and use-cases in the future

Why is this interesting for you?

• Access to privileged commercial data
  (search and click-through data)
• Test your IR systems with real, unsuspecting users in a live setting
  (not the same as crowdsourcing!)