



HCSNet Workshop

Using clickthrough data to determine relevance of search results

Bodo Billerbeck
Friday, 22nd of September 2006

Overview

- ARC hypothetical
- Opportunity for search engine companies
- SIGIR paper on clickthrough

ARC - Special Research Initiative: Hypothetical

- Area of research:
 - Research background is the efficiency of query expansion
 - Professionally, anything that enhances relevance of search results
- Key next generation search problem:
 - Deducing implied relevance from clicks (of course!)
 - Query quality
 - Personalisation (query context)

Opportunity

- Accumulated a large amount of data
- How can we improve search?
- Users are lazy, it is difficult to get explicit feedback
- Using implicit feedback in the form of clickthrough:
 - What is a reliable way for us to determine what a click means – is it random, a mistake, or because the user is convinced that the result was relevant to them?
- E. Agichtein, E. Brill, S. Dumais, R. Ragno *Learning User Interaction Models for Predicting Web Search Result Preferences*, Conference on Research and Development in Information Retrieval (SIGIR'06), 2006

Average distribution of clicks

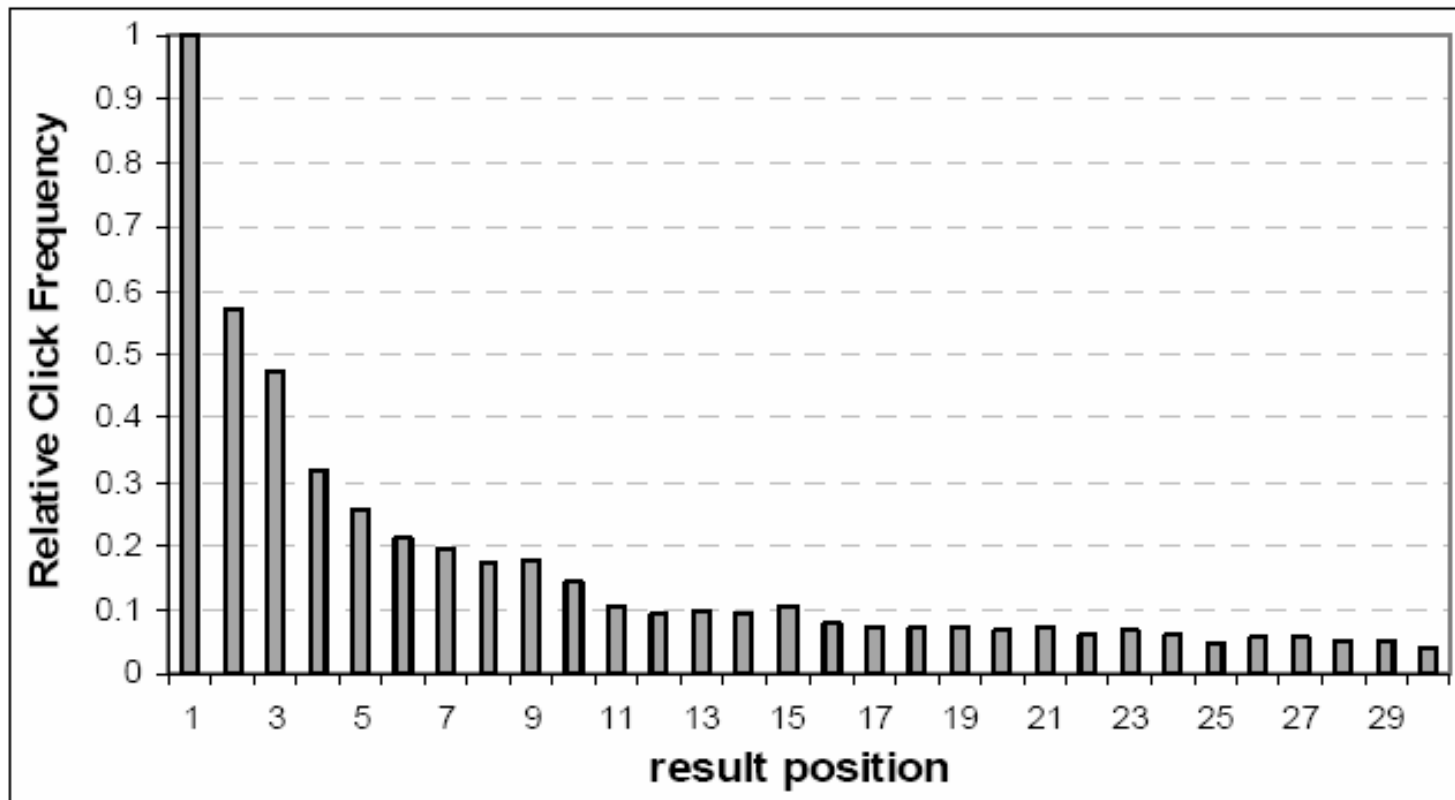


Figure 3.1: Relative click frequency for top 30 result positions over 3,500 queries and 120,000 searches.

Clicks, relative to top relevant document

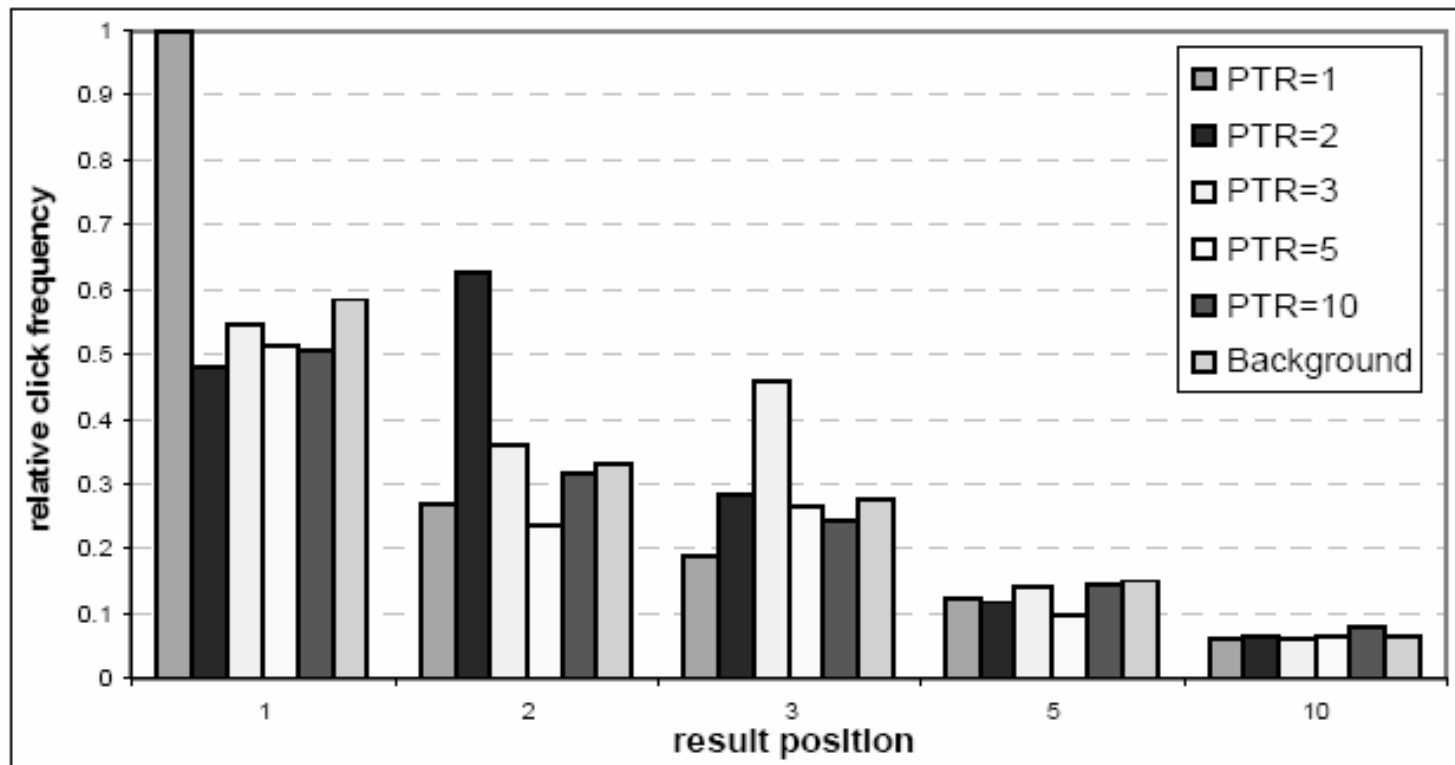


Figure 3.2: Relative click frequency for queries with varying PTR (Position of Top Relevant document).

Normalised clicks

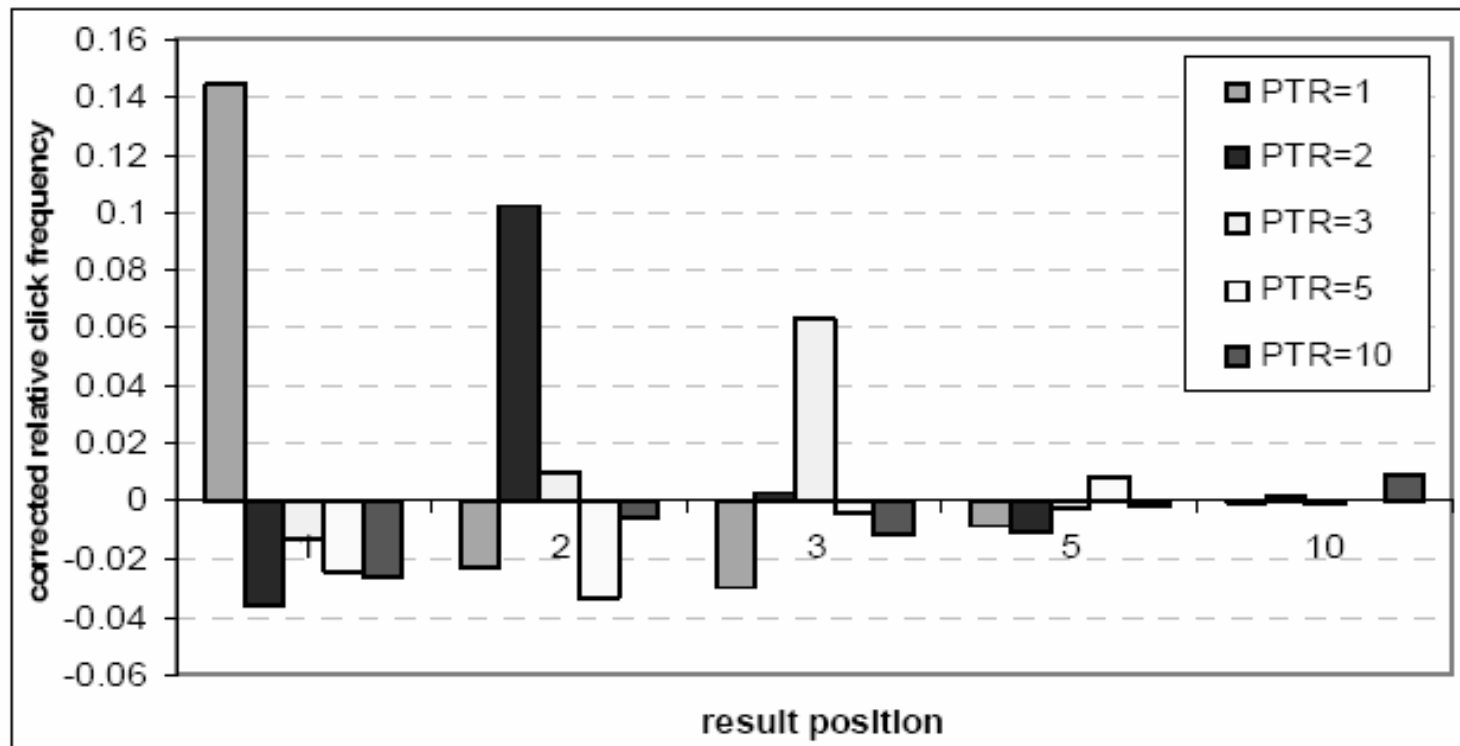


Figure 3.3: Relative corrected click frequency for relevant documents with varying PTR (Position of Top Relevant).

Problems

- Accumulated a large amount of data
 - But is it enough
- How about better formulated queries, which may not have been posed before:
 - Is it possible to learn from other past queries (that may or may not be related)?