Collecting Web Search Quality Judgements

Timothy Jones

A common way to evaluate web search is by pooling judgements. The top results from the search rankings being tested are combined into a pool. Users then judge how relevant each result is for the query it was retrieved for. This poster presents a tool we have built for retrieving pooling judgements from users.

Candidate search results are presented one by one at the top of the evaluation tool. Order of results is randomised, and results are initially displayed outside of the ranked list. In this way, order-of-presentation bias is avoided.

Users rate results using the rating buttons. In this way, a strong ordering is collected.

Once a result is rated, it is placed in the ranked list below.

Before submitting their judgements, and at any time during the rating process, users can:

- Re-rate results using these buttons
- Re-order results using these buttons

In this way, absolute judgements (this document is good) and preference judgements (document A is better than document B) are collected.

Allowing reordering of results that have been given the same absolute judgement also allows collection of weak ordering information (document A is a little bit better than B)