Making Sense of Search Result Pages

[Position Paper]

Jan O. Pedersen
Yahoo! Inc.
2811 Mission College Blvd
Santa Clara, CA 95054
jpederse@yahoo-inc.com

1. OVERVIEW

Search engine result pages are presented hundreds of millions of times a day, yet it is not well understood what makes a particular page better from a consumer's perspective. For example, search engines spend large amounts of capital to make search-page loading latencies low, but how fast is fast enough or why fast is better is largely a subject of anecdote. Another example; search engines deploy large teams to continuously improve ranking quality, yet ranking quality does not fully characterize what is valued by the consumer. To see this, consider a feature referred to as *site collapse*, the custom of clustering results from the same site, subresults indented below the main (most relevant) result. Common measures of ranking quality, such as discounted cumulative gain (DCG) [1], are not optimized by this feature, a seeming contradiction.

Much of the contradiction comes from imposing a optimization criterion that does not account for perceptual phenomena. Users rapidly scan search result pages and often make a decision as to what action to take next within a few seconds. Presentations optimized for easy consumption and efficient scanning will be perceived as more relevant even if the content is identical to (or worse than) other, more awkward displays.

Eye-tracking experiments have shed considerable light on these issues [2]. It has been noted that users consume content from top to bottom cycling through quick scan-cull-decide cycles. Presentation details are extremely determinant of whether a consumer will notice a result. For example, the presence or absence of bolding can impact click rates significantly. Once a result is noticed, the decision to click is driven by ancillary information in the result summary, such as the contextual snippets and the url.

Search result pages are becoming more complex; they can no longer be characterized as a ranked list of homogeneously presented results. It is quite common to see content modules, such as the Yahoo! movie direct display, inserted above, below or between results. The result presentations are sometimes enriched with additional data, such as quick links for homepages and in-line players for video content. In addition, interventions in the process of constructing a query are possible via search assistance technologies such as Yahoo! Gossip.

As the search industry moves forward, a deeper understanding of user searching behavior will be required to better shape increasing complex products. The industry is fortunate to have an abundance of user-behavior data; every aspect of the search experience is carefully metered. However, this data is very complex in structure and captures only a slice of Internet activity. An effective model of user interaction with the search result page, validated through qualitative surveys and eye tracking experiments will be necessary to make full use of this data.

2. RESEARCH TOPICS

Yahoo! currently conducts research into user search behavior through qualitative ethnographic studies, eye tracking experiments, analysis of query log data and live experiments. Unfortunately, the largest scale, and hence richest, data source — the log data — is the hardest to interpret since it is a partial, retrospective record of user actions. In particular, simple, aggregate click-through data is not very diagnostic because it averages over a huge range of phenomena. Insights from user modeling are critical to both developing finer-grained, more revealing, user feedback measures and in developing tracking metrics that allow Yahoo! to monitor the health of its services.

The process Yahoo! search uses to design, validate, and optimize a new search feature includes the following steps:

1. Propose a design and build a mock-up or prototype
2. Translate the stated goals of the feature into expected user behaviors
3. Conduct a usability test of the proposed feature; often this includes a eye tracking experiment
4. Validate if the design achieves the desired goals; if not iterate
5. Develop proxy measures for the desired behaviors that can be measured in the user feedback logs

6. Conduct an online test of the feature to quantify effects

7. After launch, track the health of the feature through a feedback metric

User modeling is critical for steps 2 and 5 and, of course, reflects strongly on step 1. The next few sections describes some of the the data and methods used to develop the Yahoo! search user model.

2.1 Session Analysis

One rich source of insights into consumer search behavior is query log session analysis. A session is a sequence of user actions, including query reformulations and url clicks, over a relatively short span of time associated with the same user goal. For example, one finds in the Yahoo query logs the following session fragments:

Lead query	Follow query
samsung lcd tv	best buy store
samsung lcd tv	consumer reports
37 inch lcd tv	best 37 inch lcd tv
shelving units	costco wholesale
shelving units	home library

The user's intention is made clearer in the reformulation. Joining this data with user clicks actions is even more informative. Yahoo! actively mines session data to build search assists, spell correct queries, and to automatically expand queries (in sponsored search). A newer area is mining this data to measure the effectiveness of ranking and presentation algorithms, with the underlying assumption that better selection and ranking algorithms will anticipate user reformulations. For example, one study examines conditions under which users switch from one search engine to another [3]. However, the surface has only been scratched in the analysis of session data.

2.2 Toolbar data

Search engine query logs only reflect a small slice of user behavior — actions taken on the search results page. A more complete picture would include the entire click stream; search result page clicks as well as offsite follow-on actions. This sort of data is available from a subset of toolbar users — those that opt into having their click stream tracked. Yahoo! has just begun to collect this sort of data, although competing search engines have collected it for some time. We expect to derive much better indicators of user satisfaction by consider the actions post click. For example, if the user exits the clicked-through page rapidly then one can infer that the information need was not satisfied by that page. User satisfaction indicators can then be correlated with search result page context to shed light on usefulness.

2.3 Eye Tracking

Ultimately a direct measurement of what the user perceives on a search results page would be most useful. A closely proxy of this ideal is measuring eye movement and fixations across the page followed by user interviews. Aggregate measures reveal which page elements get the greatest attention, while tracking individual sessions is very revealing about scanning behavior. One challenge is scaling up these experiments, which are quite labor intensive, so that they can be used not only to generate hypotheses but also to statistically confirm them.

3. CONCLUSION

We are still very early in the development of a comprehensive model for user interaction with the search results page. A this time, we use operating principles, such as the hypothesis that users approach a search task with a fixed cognitive effort budget or that bolding is critical is gaining user attention. We will need more quantitative models to take the search experience to the next level.

4. REFERENCES

[1] Kalervo Järvelin and Jaana Kekäläinen, *IR evaluation methods for retrieving highly relevant documents*, SIGIR 2000, pages 41–48

[2] A.T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag, New York, Inc., 2003

[3] Yun-Fang Juan and Chi-Chao Chang, *An Analysis of Search Engine Switching Behavior using Click Streams*, WWW 2005, Chiba Japan, pages 1050–1051