

How do you know your site search is working?

For a while now lots of people have been telling me that our site search is Not Much Good. We've recently got together a team of people to work on fixing it, but before we could do that we had to work out what exactly is wrong with it. All the feedback was in the form of anecdotes such as "I tried to find x... I know it's there, but I couldn't find it", "People can't find what they're looking for", "We've written something about x, but no-one ever finds it".

All worth investigating, but all quite difficult to reproduce, and even more difficult to work out whether we've fixed them.

In this article I'm going to describe some measurements we're using to help understand what's wrong with our current search, and then to be able to work out whether our improved search is better. Before that, though, you need to know a little bit about our search and about what counts as "working well" as far as we're concerned.

How our search should work

The search on our website is basically a fairly simple best-first search; that is, we want to put the best result for people at the top of the list. If it's working well, users should be able to enter their search term, find a good result easily, and they'll then leave our site from the page that has the information they need on it. We make the assumption that once they're confident they've found what they need, they won't feel they need to browse our site further.

The following characteristics need to be in place:

1. **The best result is near the top of the list** (in the top 3)
2. **Results are all relevant to the search term** (if not, users lose confidence in the search)
3. **Results are presented well**, so that users can easily scan for the best result
4. **Good results are returned the first time the user searches**
5. **Users find what they're looking for**. This is actually our main goal – all the other items in this list are factors that we believe contribute to this being successful

The search also needs to work – i.e. it needs to index the right content, searches need to return results in a reasonable time, and so on. I'm not going to be talking about how we measure this in this article, but that doesn't mean it's not important.

So... how do we measure our search?

As the list above indicates, search is a complex problem. The search engine has to work technically, the UI has to be right, the website's content needs to suit whatever users are looking for... So we measure our search by looking at a set of factors, and looking at them in combination.

1. Does the search put useful results at the top of the list?

Following the recommendation in John Ferrara's article on *A list apart*, we identify what the ideal result should be for the top 25 search terms on the site, and then find out how far down the results list this preferred result appears. It should be comfortably within the top 3 results on this list – because that's as far as the majority of search users will look. We work

out the average position of preferred results for all 25 search terms, and this gives us our “relevancy” score.

This is a manual exercise, involving the people who create the website content, along with the people who understand our site users best (e.g. our product support team).

2. Does the search return only relevant results?

The results that the search returns need to be relevant to the term they’ve entered. If they’re not relevant, then the user loses confidence in the search. This is particularly important where the preferred result is not right at the top of the results list. As a paper on search by researchers from Yahoo and Google identifies: “the likelihood a user examines the document at rank i is dependent on how satisfied the user was with previously observed documents in the ranked list”. In other words: show irrelevant results and the user’s going to be less willing to scan further down the list.

Following the methodology outlined in the *A list apart* article again, we run each of the top 25 searches and we work out whether each result on the first page of results is relevant, close or completely irrelevant. This gives us an overall percentage score for “precision”, and we aim to achieve at least 75% for this.

As with the “relevancy” measurement outlined above, this is a manual exercise.

3. Does the results display enable users to scan for good results

It’s difficult to measure how scannable the results on search pages are, because of the complexity around setting up realistic user tests here. We know we’re not currently following best practice for presenting content to support scanning, though, so we’ll work on implementing this.

4. Do the terms people use give them the results they’re looking for?

We want search users to find a useful result when they first search; they shouldn’t need to refine their search to get a good result. We’re trying to avoid “thrashing” behaviour – where users refine search over and over again, trying to find the search terms that return the results they’re looking for. This is frustrating for the user and often unsuccessful too.

We use the *search refinements* data provided by Google Analytics to measure the percentage of searches which are refined (i.e. the number of times the user searched again, immediately after performing a search). We want this number to decrease ultimately, although we’re not entirely sure yet how low we can reasonably expect it to go.

5. Do users find what they’re looking for?

We only want the characteristics we measure in #1-4 to work because we believe they contribute to the ultimate goal of search: enabling users to find things.

The search experience we want visitors to our site to have is that they perform a search, click on a result, and are so happy with what they see there that they leave the site. They don’t come back to the list of search results to see if there’s anything better (they don’t need to: they’ve already seen something that answers their question).

Ideally, we’d like to follow where people go from search results pages, and work out how many of them leave the site from a relevant page ... but that’s too complex to do on a large scale.

However, we *can* measure the other side of this behaviour (i.e. the behaviour we *don’t* want to see): the number of people who exit our site from a search results page. We use the *search exit* data from Google Analytics for this, and we want this number to be low. As

with search refinements, we're not sure how low it's reasonable to expect this to get right now... so for the moment we're just looking for a decrease.

Putting it into practice

The measurements I've outlined here are all fairly simple... but it turns out that the practice isn't so straightforward. As I mentioned previously, search is a complex problem; there's an interaction between the different factors we're measuring which we need to take into account when we analyse our data.

For example, getting the format of results right only really helps if the search engine is returning good results.

And it doesn't matter how great the search is if there just isn't any content that matches the search terms – e.g. the content uses the wrong terminology or just doesn't exist. How do you tell whether a problem with search not returning results is due to the engine or the content? (That's not a rhetorical question, by the way: how *do you* work this out?) However, we are pretty sure that with this set of measures we at least have a good starting point for our investigations into what's wrong with the search, as well as ongoing monitoring as new content is added and new search terms become common. The relevancy and precision measures (#1 and #2) even give us a way to measure our search offline, so we can know a lot about its performance before we release.

How do *you* know your site search is working?

I'd love to hear how other people measure and tune their search. What do you measure? What processes are in place for monitoring its continued success? How often do you check up on it?

Source: <http://communicationcloud.wordpress.com/2010/01/29/how-do-you-know-your-site-search-is-working/>