

Web Measurement and Analysis for Internal Search

By Gary Angel

Web analytics is a hot topic these days. The demand for Web analysis is skyrocketing. And chances are that your organization has made a significant investment in enterprise Web analytics software sometime in the last year or two. But what exactly is Web analytics? You can read a marketing boilerplate about measuring drop-off and studying paths till your eyes cloud over - but it won't tell you much about how Web analytics is actually done or give you a good working sense of what it can do. In this article, I'll walk through a series of steps for analyzing one of the major components of many sites - internal search. I'll step through the types of analysis an analyst needs to do, explain why they're important and what you might hope to discover by doing them. In doing this, I have two purposes. One goal is to help anyone genuinely interested in analyzing internal search do a better job. The second reason is to give marketing and Web managers a much richer sense of what Web analytics is and how it actually works.

Thinking about Search

For many sites, internal search is the single most important navigational tool for users. Search is especially important on large publishing sites with lots of archival information and for retailers with a large number of products. For these sites, search can be found in more than half of all visits - and can dwarf even the home page in importance. What's more, almost all retail sites find that search is more effective as a tool for converting visitors than traditional directory navigation.

But not all sites rely on search. Many smaller sites don't support search functionality. And for large sites with a narrow focus, search is used more as a fallback - an option for users who can't find what they are looking for.

What makes this distinction (search as a preferred navigation option versus search as a fallback option) important is that it affects the types of analysis that need to be performed. In the first scenario, the goal is to find ways to maximize search results, usage and navigation to increase conversion. But in the second scenario, much of the analysis is centered on keeping visitors from having to resort to search at all.

What makes search different than any other page?

- Search changes with nearly every invocation - analyzing link paths from search is tricky.
- Search is accessible from every page on your site.
- Search has important additional data items - the search term entered and number of results.
- The search results page is often "refreshed" as visitors click on next pages of results.

One other special note - since search is most often a text box on every page, analysis will typically be looking at the search results page. The search results page can be tracked in most tools requested pages, pathing and from/next step reports.

Getting a Handle on How Search is Tracked

In Web analytics, you often need to do some basic homework before diving into the numbers. And nowhere is this truer than search. Your first step should be to make sure you understand how the search results page is being captured in your tool. Here is a checklist to use when looking at your search results page and your Web analytics solution:

Check	Method
Page Name	Check to see if a new Search, an advanced Search, and browsing to the next page of results trigger different page names in the tag. When the name doesn't change, most tools will record the action as Page Refresh .
Search Term	Check to insure that Search Terms are being passed and populating.
Search Results Count	Check to make sure this shows up in the report suite. Check to see if searches that return zero results are identified as failures.

When evaluating search usage, always keep in mind how your site and measurement tool are treating each key behavior - browsing additional pages of results, advanced search, and repeat searches may result in distinct page views or, more commonly, refresh views. It's absolutely vital that you understand the difference.

If you are in the position of setting up tags for internal search, then it's best to make sure that each of these behaviors does result in a distinct page name/page view - both for basic and advanced search results.

Evaluating the Importance of Search

There's a good chance you already know whether search is a primary navigation tool or merely a fallback mechanism on your site. Still, it never hurts to have real numbers to back up those intuitions.

There are several ways to think about tool usage - the most common being how many page views does it get. For internal search, your basic page view counts are often misleading because of the problems around refresh views. And because search is essentially a navigational tool, what's most important is how often it is used by session and by visitor.

Almost every tool will give you the views and visits for search (and for your site) right out of the box. Getting a visitor usage rate may be trickier. Your Web measurement tool will likely provide you a monthly uniques count for the site as a whole. It may or not provide you with a good monthly uniques count for the search results Page - forcing you to segment by search usage to get a visitor count.

When you have the usage rates (search visits/site visits), you'll have a good indication of how search fits into your current site design. Typically, sites that treat search as a fallback option will show visit usage well under 10 percent with three to five percent being typical. Mixed-mode sites typically range between five to 20 percent - with the higher the number the more clearly search is a preferred navigation option. Above 20 percent and search is clearly one of - and probably the - most important element on the site.

Even when only a small percentage of visitors use search, it may turn out that for those users it is the first thing they do on your site. If everyone thinks of search as a last resort on your site, this statistic can definitely raise eyebrows and is often used to help build a case for getting a better internal search System. To find out if this is true, you can study when users visited search in the Session.

One of the simplest but least informative ways to get the when is to use the average page depth statistic. While occasionally revealing, this number provides no distribution information and - even when it happens to provide an answer that turns out to be meaningful - you won't know its right unless you do more analysis.

Usually, you'll need to do that analysis using your tool's event or path reports. Pathing is often trimmed data or very difficult to consolidate; for this type of analysis, you don't need big chunks of data. If your site generates lots of search activity, you can restrict this analysis to a single day.

Even on fallback sites, 30 to 60 percent of search perceptions about when search is used and how important it is to have a good internal search system.

Search Sourcing - What Gets Visitors to Search

Traffic to most pages is from a few dominant sources - highly related pages or major navigational elements (like top-nav). Search is different. It will be sourced by lots of different pages, and for tools that are available globally, one of your first analytic directions will be to understand places that are heavy or light sources of traffic.

You can get the raw counts of search Sourcing in most tools by using the previous pages (or pages from) reports. This will give you the information about aggregate sourcing. It won't tell you how each of these sources compares in terms of their rate of sourcing. Rate of sourcing is a simple calculation - search sources/views of source page. Why should you care about the rate? For sites where search is a fallback mechanism, resorting to search is an indication that the content doesn't provide the information a visitor is seeking.

Conversely, where search is a primary navigation device, spotting pages where the search Rrte is low can help you identify content where search is being underutilized.

To get the rate of searching, you just need to match up the sourcing counts (next page is search) to the page view counts for each source. In other words, if a page sources 60 searches and has 6,100 views, then its search rate is just under 1 percent.

So how do you use this rate? Establish an average search rate for the site and compare your actual page search rates to the average.

For fallback sites, if a page is search sourcing significantly more than average then it may have a problem. For primary navigation sites, the analysis is reversed. Pages that source to search less than average may need strong search drives. Naturally, this analysis isn't applicable to every page on your site. You don't want visitors searching in the middle of an order process.

Measuring Search Performance

One of the easiest ways of measuring search performance is to start with a functional approach - measuring search by what it is supposed to accomplish.

Think about it and you'll see that search is a page designed to move visitors to the appropriate place on the site. What's more, search as a tool on your site is competing with pages that are usually classified as "router" pages - pages whose primary purpose isn't selling anything but just getting visitors to the right place.

When you measure router pages, the best key performance indicator is the percent of routes that actually go in the direction intended. These are called "body" routes because they typically exist in the main body of the page (as opposed to top, right or bottom navigation). The higher the percentage of routes directed to the target, the better the router.

Search fits a little uncomfortably into this paradigm because it isn't designed to route to any particular set of pages. However, that doesn't mean you can't look at search as a router. You just have to be a little creative about how you are going to classify routes.

Generally, you should assume that search routed to an intended destination when the "next" page isn't one of the following: a site exit, another search, the home page or to another navigation page. Why? Search is supposed to move people down into specific areas of your site - ideally to the most directive content. That means search shouldn't be taking people to router pages. Otherwise, you're just stacking up navigation options.

Using this strategy, you can break down search next pages into a few basic functional categories: site exits, backward moves, sideways moves, and body or intended routes. In each case, the value should be expressed as a percentage of total routes. Here's an example:

[figure 2]

You'd be surprised how often this simple analysis based on the classification of routes will provide you with a better understanding of the effectiveness of search.

In essence, what this method measures is how well search is driving visitors to any of its results. Those results may not be "good" results, and, at this level, your analysis is far from complete. Search can perform very well for some types of searches and poorly for others. Is there a way to break down search performance by category?

There is, but it isn't available in every tool and the exact method varies. The basic idea is simple: create a segment based on visitors who "searched" specific types of terms. Then look at the search performance for just those visitors/visits.

Evaluating Internal Search Performance by Topic

It is in the nature of search to perform differently depending on the search term or concept. So if you are doing a real deep-dive analysis of search, it's essential to evaluate search performance for the key types of queries on your site.

To do this, first isolate the search traffic generated by a specific query or set of queries. In most cases, this will involve using your tool's segmentation

capabilities. But first, you will have to find the appropriate terms. Start with a report on internal search terms.

Most Web analytic tools will provide this report - showing a list of search terms and a count of the times used. Find one or more terms that are applicable to your "key" concept. Now use your tool's visitor segmentation tool to create a segment that entered one or more of these concept terms.

Next, apply the segment and track your routing performance just as you did for all of search. It's interesting to compare the routing performance of categories to each other and to search as a whole. When you've isolated a single concept in search, and can compare the performance of search to both the corresponding router page and other concepts, then you can really evaluate how well search is working.

This concept-based analysis of search may turn up very considerable tuning opportunities. Often, the most effective search optimization is in customizing the output in response to specific terms. This takes real work, but the impact of customizing your top search results can be higher than what you'll get by replacing your search engine with a significantly better one.

Like any type of analysis, Web measurement is most productive when approached in a disciplined fashion. Many analysis tasks follow a similar path: how do the pages on the Web site register in the measurement solution, how often/intensively is the page or tool used, how and when do visitors access the tool and how effective is it in the performance of its function. These questions are equivalent to a journalists "who, what, when, where, why" framework. Translating them into analytic reports is not always simple - and the techniques covered for internal search range from quite basic to very sophisticated. But taken together, they provide a path for really using Web analytics to improve your internal search - and, by example, many other parts of your Web site as well!