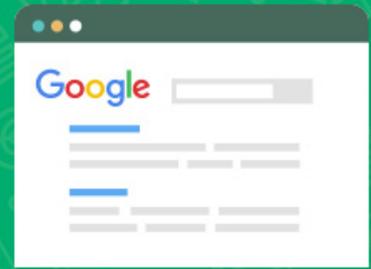


1 Million Results

Search Engine Ranking Factors Study



BACKLINKO RANKING FACTORS STUDY:

Methods & Results

BACKLINKO

BACKLINKO RANKING FACTORS STUDY: METHODS & RESULTS

To our knowledge, this is the largest analysis of Google's search results ever conducted.

Our four data partners for this study were [SEMRush](#), [Ahrefs](#), [SimilarWeb](#), and [MarketMuse](#). We also looked at 1 million unique webpages via [Alexa's API](#). Additionally, we built our own application to crawl 1 million webpages for many onsite ranking factors (for example, a page's word count).

We decided to conduct this study in order to better understand the relationship between various proposed Google ranking factors and actual Google rankings.

**To understand these relationships we analyzed
1 million unique Google search results.**

The largest search engine ranking factor studies we know of for were conducted by [Moz](#) and [SearchMetrics](#). Both of these were extremely well-done and helped inform how our study was conducted.

Most previous search engine ranking factor studies used a dataset of approximately 10,000 keywords. While 10,000 keywords is a significant sample, it doesn't begin to capture the breadth of the [millions of unique Google searches](#) performed every day.

To get a better picture of which factors influence rankings across a wide range of keywords and industries, we decided to analyze 1 million results.

WHAT WE DID— STUDY METHODOLOGY

We took 1 million random keywords from SEMRush’s database of 80 million (MM) Google search engine result pages (SERP). Their database contains a wide range of keywords (“keyword” is used to mean one or more words in a search query). This includes extremely popular keywords all the way down to those with as little as 5 searches per month.

From the one million results, we looked at the top ten ranking URLs for each keyword (10 million total). From that data set we pulled a random subset to examine 1 million unique URLs (the total URLs from the 10 million were 6.5 MM unique URLs).

By taking a random 1MM URLs from the 10MM we were able to achieve a data set that contained a diverse set of unique web pages. Therefore this study should apply to internet marketers working on or at websites of all sizes.

It’s important to note that Wikipedia, Facebook, Twitter and the other large websites have a larger effect on some metrics vs. others. This is especially the case where top websites have standards sitewide. HTTPS is the best example.

To calculate correlations we used [Spearman Correlation](#) for all measures. This correlation is most appropriate when looking at how one or more variables effects rank positions. Like any correlation study, correlation does not always indicate causation. Therefore we were very cautious about interpreting the results from the study.

OUR FOCUS ON THE TOP 10 RANKING PAGES

By focusing on the top 10 ranking pages in Google, we focused on we called the “winner’s circle”. On average, 71.33% of searches result in a page one organic click. Page two and three get only 5.59% of the clicks. On the first page alone, the first 5 results account for 67.60% of all the clicks and the results from 6 to 10 account for only 3.73%. These figures are from Philip Petrescu’s results published [here](#).

In other words, we wanted to find data that would help a website move from a #10 ranking to the top 3 results. Of course, our results also apply and extrapolate to pages 2, 3, etc. However, our results are looking at a narrow band of first page results. By confining results to top 10 we are looking at the difference between the rankings that matter most.

Other rank tracking studies looked deeper into SERPs. This means their correlations have a wider variance. With the wider range of 30 “ranking” scores, we’d expect the correlation will be stronger with ranking of

1-30 vs 1-10. There is less variance in the 1-10, and undoubtedly they have more in common with each other (including that fact that we’d expect to find far more from the study group among them).

When we show a correlation of .07 it suggests a ranking signal that is twice as powerful as another ranking signal measure of .035. Think of the massive size of 1 MM unique ranking webpages in our sample. While many correlation studies stop at .05 for relevance, we are confident down to .03. Below that number we label our finding as little or no correlation. Clearly, .035 is a very small relationship. However, when you’re just looking at the “winner’s circle”, a few of these small differences can mean the difference between a page ranking #1 vs. #7.

THE BACKLINKO CRAWLER

We built a script to crawl the IMM unique pages in our data set. This crawler pulled the fully rendered web page, which includes javascript and raw HTML. We primarily used this crawler to look at factors including title tag, H1, number of words per page, HTTPS vs. HTTP, YouTube video tags, use of schema markup, and image tags. We successfully pulled data for all pages excluding PDFs, Word docs, and other very uncommon document types. For all factors, we succeeded with over 98% of pages, excluding the aforementioned file types.



Google announced in May, 2014 that [it is in fact rendering JavaScript](#). When one tries to “score” a page for SEO audit keep in mind that most advanced SEO tools on the market cannot render JavaScript. Our study results fully take into account the content Google sees from rendering JavaScript. One of the most important places this impacts our results is the number or words in the body text. For example, a page could have 500 words without rendering javascript, but 800 words after rendering the page as users and Google would see it.

DATA PARTNERS AND PROVIDERS

SEMRUSH

Google ranking information is updated on a rolling basis, with some words updated hourly and some once a month. We did not look at Bing or Yahoo for the study. We extracted the real time data in the first week of December, 2015. The Ahrefs data was pulled at the end of December.

We looked at SERP data for non-mobile searches on desktops (non-mobile). SEMrush SERP results are USA, logged from a national level. So, no there is no local bias in the data set.

Non universal search items in SERP (for example, video, maps, and knowledge graph results) are effectively purged by SEMrush when determining rank numbers.

By volume of data analyzed, **Ahrefs** was as large as SEMrush. All our metrics from their data is from our set of 1 million unique URLs.

Our data from **SimilarWeb** was from a sub-set of 100,000 unique domains. Their scores are on a domain and subdomain level. In other words, we matched SimilarWeb's domain-level data to the pages in our set of 1MM unique webpages. Obviously, there is variance from one page on a domain to another. However,

the average bounce rate for the entire domain shows a fairly strong correlation with ranking. As with many metrics, one has to question if another factor is responsible for the correlation and whether they "help each other." For example, if a page or website has thorough semantic topic coverage, that might assure a searcher's quest for a full answer is covered and bounce rate is low. As this is a correlational study, we're not able to assign causation to any factor.

We used **Alexa** to collect page load speed data.

Our "comprehensive and thorough topic coverage" comes from **MarketMuse**, a leader in use of semantic algorithms to score content. To our knowledge, no other study used an application with this level of sophistication to score semantic content.

BREAKDOWN BY RANKING MEASURE

SEMRUSH

Our measures relied on their ranking data of keywords, URLs, and rank. They are the undisputed most comprehensive and accurate data provider for organic ranking.

AHREFS

AHREFS DOMAIN RATING

We looked at Ahrefs domain rating for each URL. Some domains have multiple URLs associated, others do not. [Here is an explanation](#) of Ahref's Domain Rating: "When Website 'A' has a ranking and website 'B' is linked to it, then more ranking will pass to the Website 'B' domain. Similarly, the same domain with multiple outbound links will transfer low rankings to each link. Once this process runs a dozen times, a fixed number will be assigned to the website, thus creating the Ahrefs Domain Rank."

ANCHOR TEXT

We looked at up to 10 inbound links for each URL. The exact match measure was a yes/no for appearance of the entire keyword in the anchor text.

When a URL had more than one link, we used a percentage of anchors that were exact match. So if 2/3 of the links contained exact match anchor text, then the percentage was 66%. For the partial match, we derived a percentage of the keyword (keyword phrase) that is in the anchor text. So if 1/3 words match its 33% or .33.

REFERRING DOMAINS LINKING TO URL

This data came from Ahrefs.

AHREFS URL RATING

[Ahrefs defines URL rating](#) as a metric that "measures the impact of all backlinks with different link juice to a given page. Every URL has its URL.

Let's imagine the process of URL Rating building as a couple of steps or iterations:

- A downloaded page gives its rating to all pages it refers to (only dofollow links are being counted);
- The rating value such a page adds depends on its rank and the number of links it has.

For example, Page has the rating of 100 and has ten dofollow links. Page can pass no more than 80% of its rating to another page, which means Page with the rating of 100 passes the rating of 80. This rating 80 gets divided between the dofollow links; each link will get the rating of 8."

MARKETMUSE

CONTENT TOPIC AUTHORITY

The MarketMuse Keyword Relevance Engine™ finds related topics and [scores the number of keyword topics a page covers](#). In short, a page receives:

1 point for each relevant topic mentioned once

2 points for each topic mentioned 2 or more times

THE BACKLINKO CRAWLER

CONTENT CONTAINS AT LEAST 1 IMAGE

This was a simple yes or no test.

USE OF HTTPS

This was a simple yes or no test.

URL LENGTH

Number of characters in each URL. This is the length after the “.com” - so the measure does not include the domain name itself.

KEYWORD APPEARS IN TITLE TAG (EXACT MATCH)

This was a simple yes or no test.

PRESENCE OF SCHEMA MARKUP

We looked for <div itemscope> AND appearance of itemtype=["http://schema.org/](http://schema.org/)

SIMILARWEB

BOUNCE RATE

A lower percent means less bounces on average.

TIME ON SITE

This was measured in seconds.

SERP CLICK THROUGH RATE (CTR)

This was measured as a percent.

RESULTS (CORRELATION DATA)

RANKING FACTOR	SPEARMAN CORRELATION
URL Length	0.13105
Use of HTTPS	0.0401
Presence of Schema Markup	0.00026
Keyword Appears in Title Tag (Exact Match)	0.021
Exact Match Anchor Text	0.051
Partial Match Anchor Text	0.01921
Content Contains At Least 1 Image	0.04253
Number of Images In Content	0.009
Content Topic Authority (MarketMuse)	0.046
Total External Backlinks	0.187
Number of Referring Domains	0.2004
Webpage Link Authority (Ahrefs URL Rating)	0.1297
Bounce Rate (SimilarWeb)	0.092704
Content Total Word Count	0.03732
Domain Link Authority (Ahrefs Domain Rating)	0.05101
Average Page Load Speed For URL's Domain (via SimilarWeb)	0.0902