

# **BACKLINKO + AHREFS GOOGLE RANKING FACTORS CORRELATION STUDY:**

Methods & Results

---

**BACKLINKO**

# BACKLINKO + AHREFS GOOGLE RANKING FACTORS CORRELATION STUDY: METHODS & RESULTS

---

To our knowledge, this is among the largest analysis of Google's search results ever conducted. Several studies (including those done by [SEMrush](#) and [SearchMetrics](#)) have looked at similar potential ranking factors. Both of these were extremely well-done and helped inform how our study was conducted.

Our main data partner for this study was the popular SEO software, [Ahrefs](#). We also received assistance for content grading from [Clearscope.io](#). Additionally, we used a scraper 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 1,183,680 keywords (1,183,628 after data cleaning with a total number of 11,835,086 ranking URLs (10,052,136 unique URLs).

# WHAT WE DID— STUDY METHODOLOGY

---

We took 1.18 million random keywords provided by Ahrefs from their database. 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 (11.8 million total).

It’s important to note that Wikipedia, Facebook, Twitter and other large websites have a larger

effect on some metrics vs. others. They also tend to block scrapers. Which is why many large sites were omitted from certain analyses.

To calculate correlations we used Polynomial regression or linear regression for all measures. This correlation is most appropriate when looking at how one or more variables may affect 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 what we called the “winner’s circle”. Less than 1% of all Google searchers go to the second page of the results. On the first page alone, the first 3 account for 75.1% of all the clicks. The source for this CTR data can be found [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 the 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 would be stronger with rankings 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).

# DATA PARTNERS AND PROVIDERS

## **ahrefs** **AHREFS**

Ahrefs provided data for Google ranking positions, the URLs ranking on those results, as well as URL Rating and Domain Rating. Ahrefs SERP tracking 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 January, 2020.

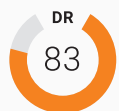
## **@Alexa** **ALEXA**

Our Time on Site and PageSpeed data from [Alexa](#) was from a subset of approx. 10,000 unique domains. Please note that their data is on a domain and subdomain level. In other words, we matched Alexa's domain-level data to the pages in our set of 11.8M unique web pages. Obviously, there is variance from one page on a domain to another. 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.

## **clearscope** **CLEARSCOPE.IO**

Our "Content Grade" comes from [Clearscope](#), a software tool designed to assist in producing comprehensive content. They provided results for a subset of 1k search results.

# WHAT WE ANALYZED —



## 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.



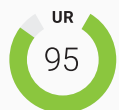
## REFERRING DOMAINS LINKING TO URL

This data came from Ahrefs.



## TOTAL BACKLINKS

This data came from Ahrefs.



## AHREFS URL RATING

[Ahrefs defines URL rating](#) as a metric that "shows the strength of a target page's backlink profile on a logarithmic scale from 0 to 100".

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



## CONTENT GRADE

The Clearscope content grader finds related words and phrases and scores the number of related subtopics a page covers.

`.com/abcde`

## 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 OR PARTIAL MATCH)

Based on whether the keyword a page ranks for appears in the title tag.



## PRESENCE OF SCHEMA MARKUP

We looked for `<div itemscope>` AND appearance of `itemtype="http://schema.org/`



## TIME ON SITE

This was measured in “minutes:seconds” as provided by Alexa.

Full breakdown of the analysis and results can be found [here](#). We also have the raw data used in this analysis [on Github](#).