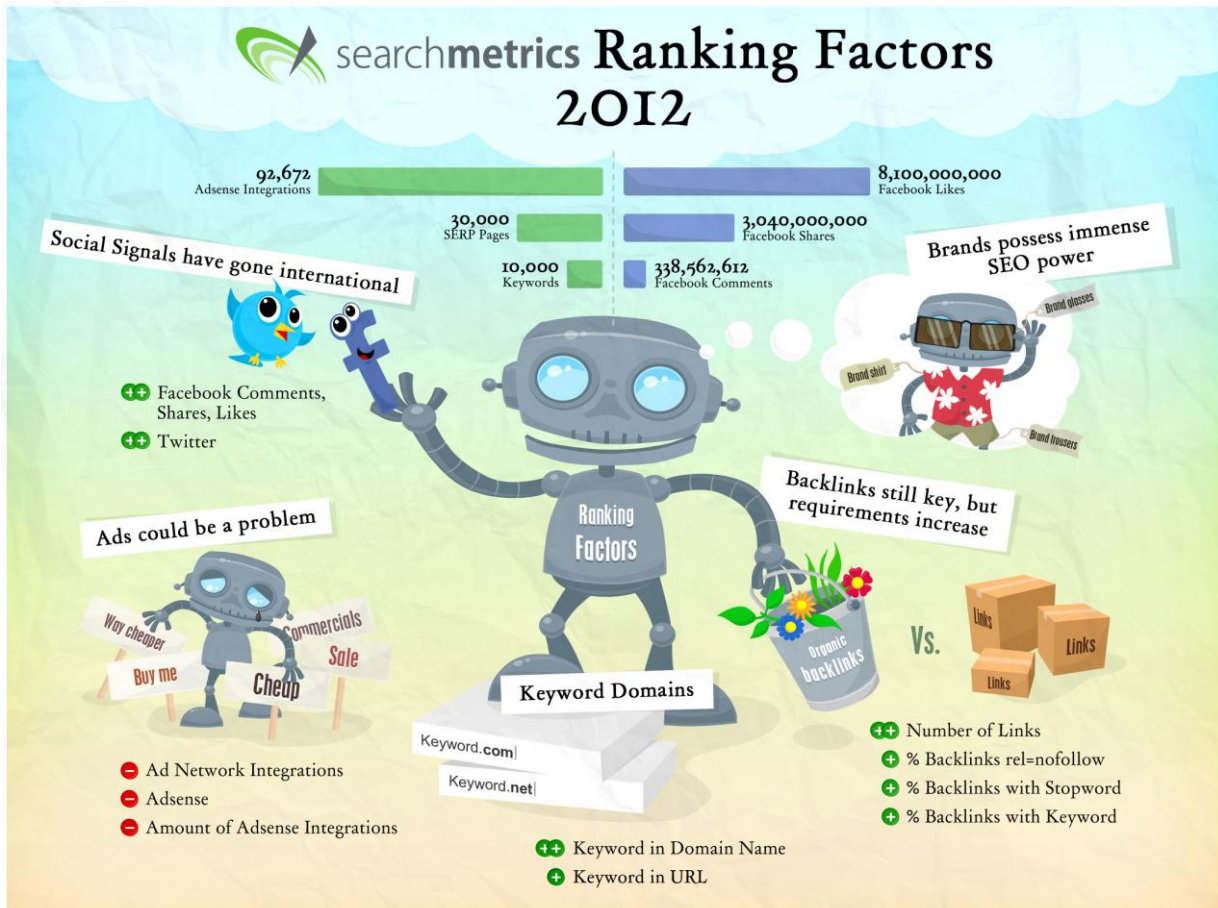


Whitepaper – U.S. SEO Ranking Factors 2012



Authors: Marcus Tober, Sebastian Weber

Searchmetrics Inc.

1115 Broadway

12th Floor, Room 1213

New York, NY 10010

Phone: 1 866-411-9494

E-Mail: sales-us@searchmetrics.com

Contents

INTRODUCTION.....	3
FACTOR OVERVIEW	4
DATA COLLECTION.....	5
THE MOST IMPORTANT FACTORS.....	5
1. SOCIAL SIGNALS ARE AS STRONG AS EVER.....	6
2. (CLUMSY) ADVERTISEMENTS ARE AN OBSTACLE.....	7
3. BACKLINKS ARE STILL SEO GOLD – BUT STANDARDS ARE RISING	9
4. BRAND POWER ENDURES	11
5. KEYWORD DOMAINS & URL KEYWORDS	13
6. ADDITIONAL ON-PAGE FACTORS.....	14

Introduction

The biggest unknown in SEO is always the algorithm. Many myths and rumors surround the what and how of Google results and their sorting and how you can best influence them.

At Searchmetrics we continually aggregate billions of data points a month and naturally look for the answer to the question: “Which factors are relevant for a good ranking in Google search results?”

With this study we will get closer to the answer on the basis of 10,000 selected top-keywords, 300,000 websites and millions of links, shares and tweets.

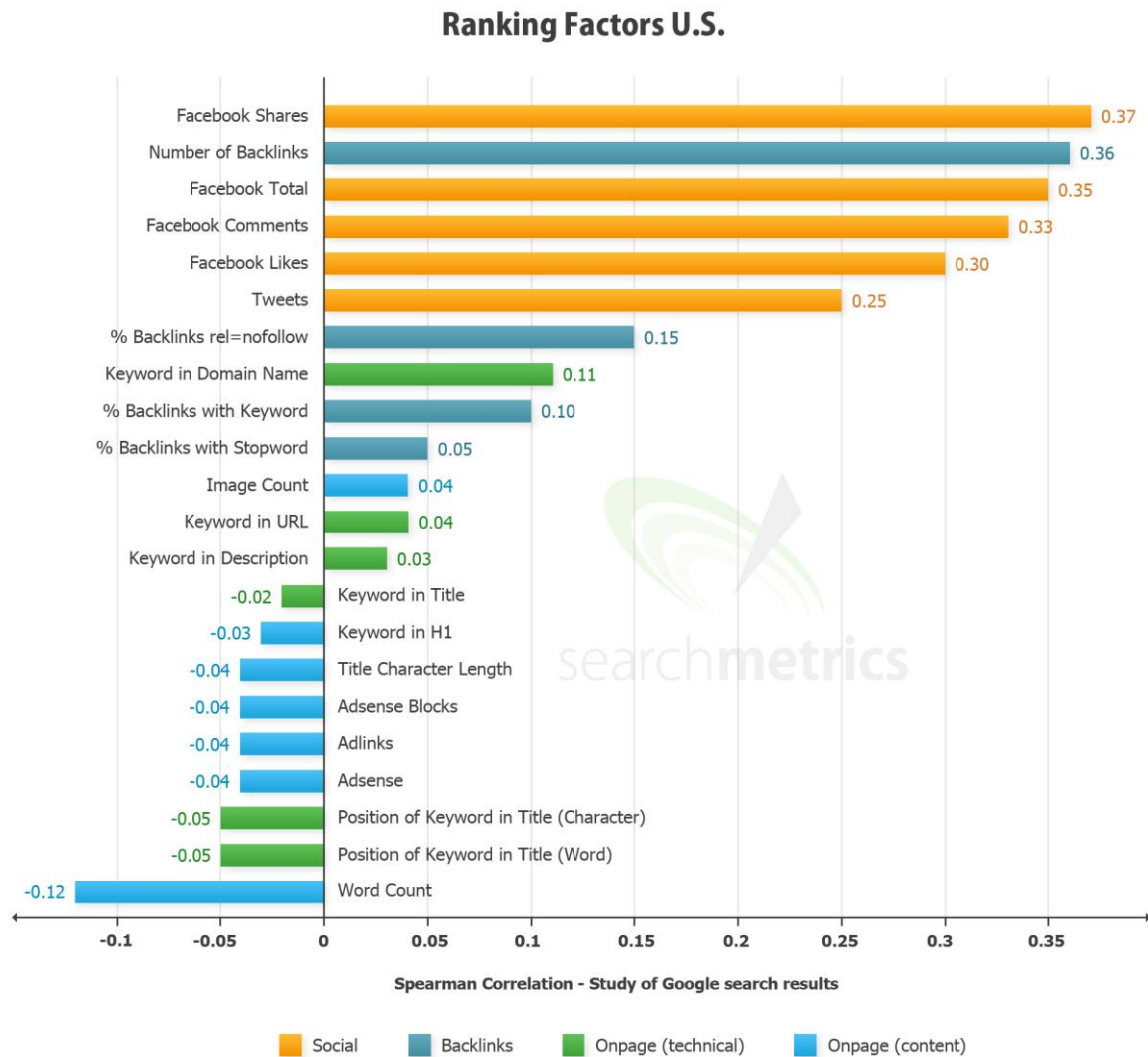
We will compare potential ranking factors and thus website characteristics with their corresponding Google rankings by assessing their statistical correlation. For example, if there are many pages in the top positions of analyzed SERPs with keyword title tags, then we have identified a high correlation.

A few highlights from our inquiry

- **Social media signals show extremely high correlation:** social signals from Facebook, Twitter and Google+ are frequently associated with good rankings in Google's index.
- **Too much advertising is detrimental:** for the first time we are seeing sites with too many advertisements struggling to rank well. However, the problem apparently relates only to AdSense adblocks...
- **Backlinks are still important but quantity is not the only important thing:** even though the number of backlinks is still the most powerful factor, links with stop words and 'nofollow' should also be included in the link-mix.
- **Brands leverage classic SEO signals:** apparently pages with strong brands do not need be as concerned with the areas of title tags, headings etc. According to our figures, this group operates under different rules.
- **Keyword domains still frequently attract top results:** despite all the rumors to the contrary, keyword domains are still alive and well and are often in the top rankings.

Factor Overview

The clearest way to present the correlations between different factors and Google search results is Spearman's rank correlation coefficient:



The larger the bar, the greater the correlation. The correlation coefficient is displayed on the x-axis. Greater values along the x-axis (e.g. Facebook Shares) have a positive correlation (the more, the better) while lower values (Title Character Length) have a negative correlation. Therefore, we can say that the largest correlation occurs between Facebook Shares and the lowest between the Position of Title Keywords. With negative values, the correlation between

title character length can be seen: the shorter the title is, the better the ranking – more about this later.

Data collection

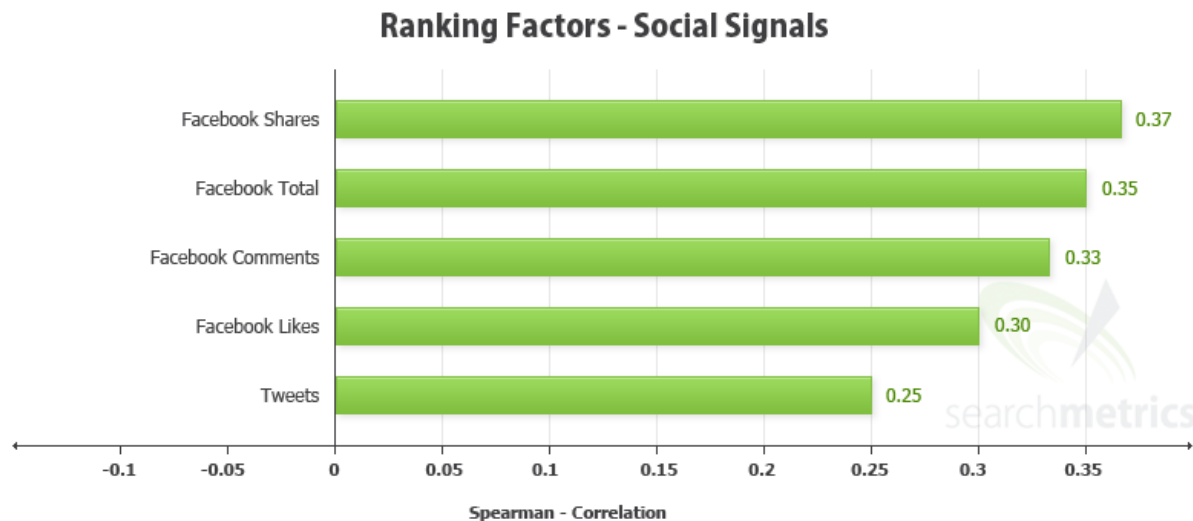
Data for this study was collected in February and March 2012. In the results, you can see the effects of the various Panda Updates that have greatly changed the look of results since the start of 2011.

The most important factors

These values are only really useful when they are organized thematically, examples given and the structures behind them ascertained. This is the goal of our study.

1. Social signals are as strong as ever

Facebook and Twitter signals correlate as follows with higher rankings:



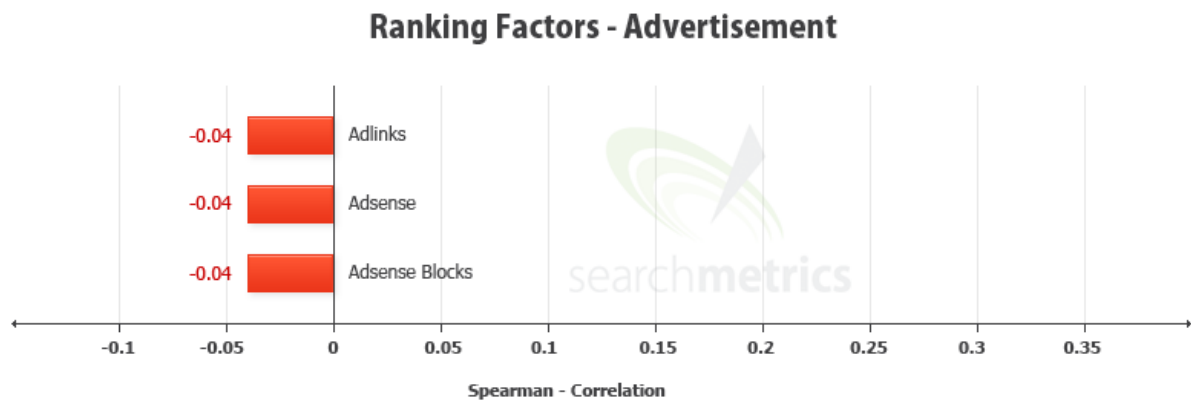
The different Facebook metrics feature the highest values, of which 'shares' appear to have the strongest association, followed directly by the number of backlinks in the overall summary. Twitter is far behind these values but is still the 6th strongest metric in our analysis behind Facebook and the number of backlinks.

A note on Google+: analyzing Google +1s with a Spearman correlation, we found a significant result of 0.41. From this we can assume that the quantity of +1s has the strongest correlation of any of the metrics analyzed in the study.

However, we have not included this figure in the overview because we consider it to be too unreliable. This is because Google+ does not currently have enough users and the possibility of a +1 leading directly to changes in SERPs follows accordingly, since pages receive +1s in the order that they would already be placed without them. When Google+ has values that are stronger and more independent from SERPs, these values will also be included in the overview. That Google is trying to make Google+ an important player is indisputable and therefore SEOs should be sure to keep an eye on further developments.

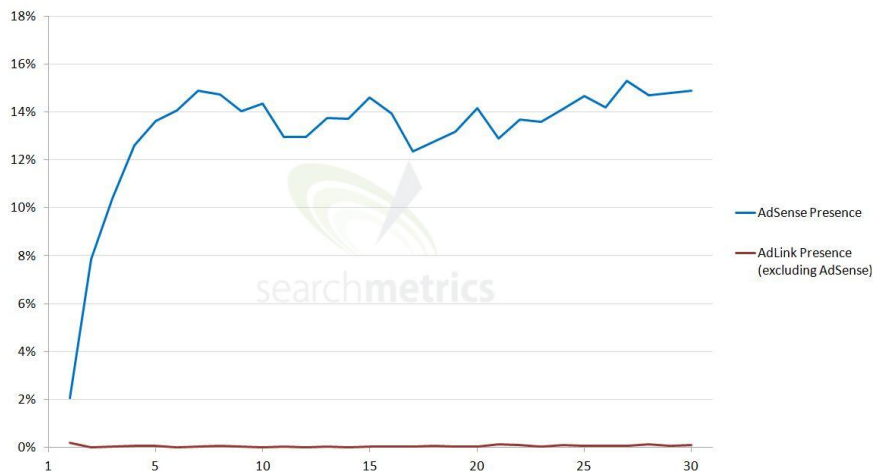
2. (Clumsy) advertisements are an obstacle

Too many and/or excessively clumsy advertisements were presumed to be a factor in the Panda Update and its successors. The data in this study support this assumption as all our analyzed advertisement factors returned a negative correlation:



General AdLinks (common integrations e.g. Commission Junction, AdSense and others) are slightly less negative than the use of AdSense alone.

However it is important to note that the correlation value above is for AdLinks for all integrations including AdSense. If we take the % trend for AdSense integration and all other analyzed competitor networks according to rankings, we arrive at a surprising conclusion:



We can clearly see that AdSense advertisements drop sharply among the top rankings.

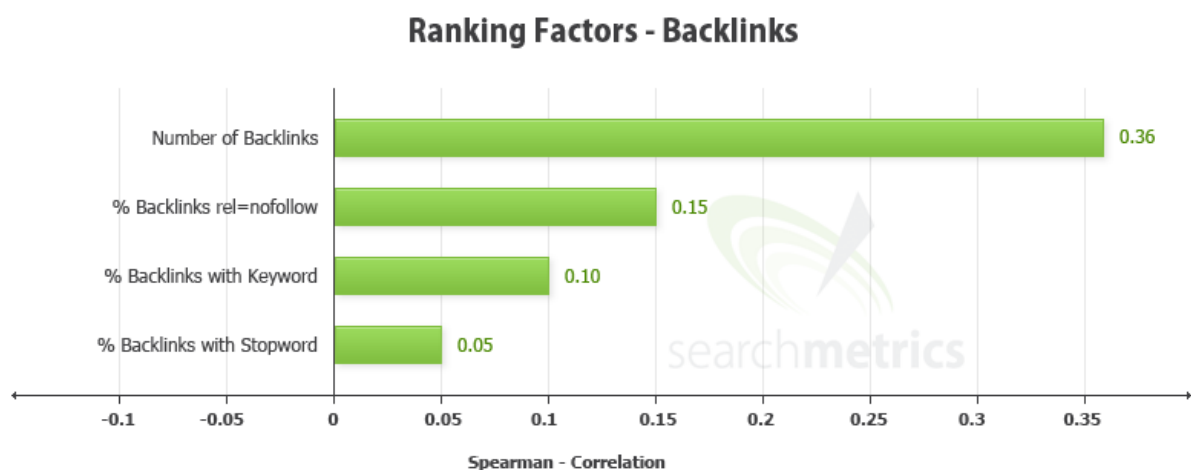
However, all other forms of advertisement that we analyzed have in fact remained consistent. The bottom line is, then, that only AdSense has a negative correlation. Therefore, at the moment it seems that we are more likely to find sites with fewer traditional advertisement integrations in top rankings. Of course, this might also have something to do with brand sites – Sony, for instance, has no banner ads on their product pages.

Recently Google even confirmed directly that particularly prominent, distracting or above-the-fold ads could lead to ranking problems – and it is precisely this trend that we can see in our correlation data. However, we continue to provide additional information since these factors only appear to be relevant in the top 10 results and because AdSense seems to feature more often on poorly ranking sites than those with better rankings.

3. Backlinks are still SEO gold – but standards are rising

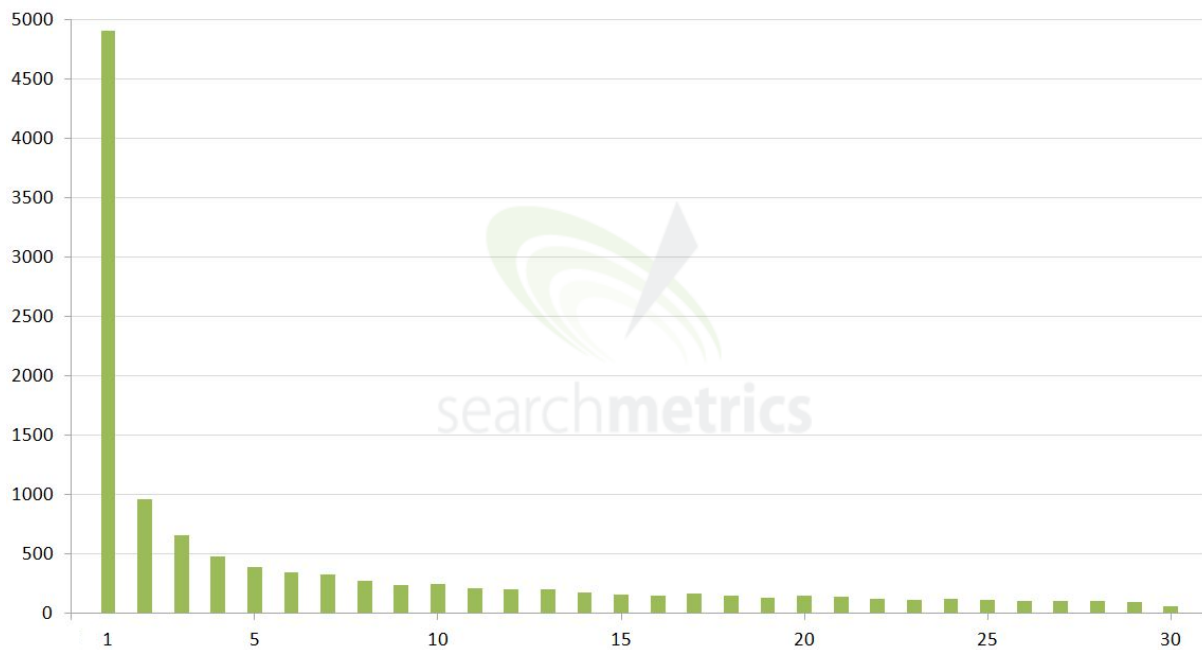
Regardless of the rising power of social media, backlinks are and will continue to be one of the most critical factors in achieving good rankings.

The correlation data supports this – following Facebook metrics, the number of backlinks is the factor that most strongly correlates with good rankings. Moreover, there still appear to be other quality factors at play when dealing with backlinks:



These figures indicate that the proportion of nofollow links correlates more strongly with rankings than the proportion of links containing keywords. Even the proportion of links containing a stop word can have an effect. This strong correlation for factors that seem to suggest a more natural link structure illustrates a trend suspected by many SEOs, that dull, perfectly keyword optimized links are often no longer effective and that another strategy is necessary.

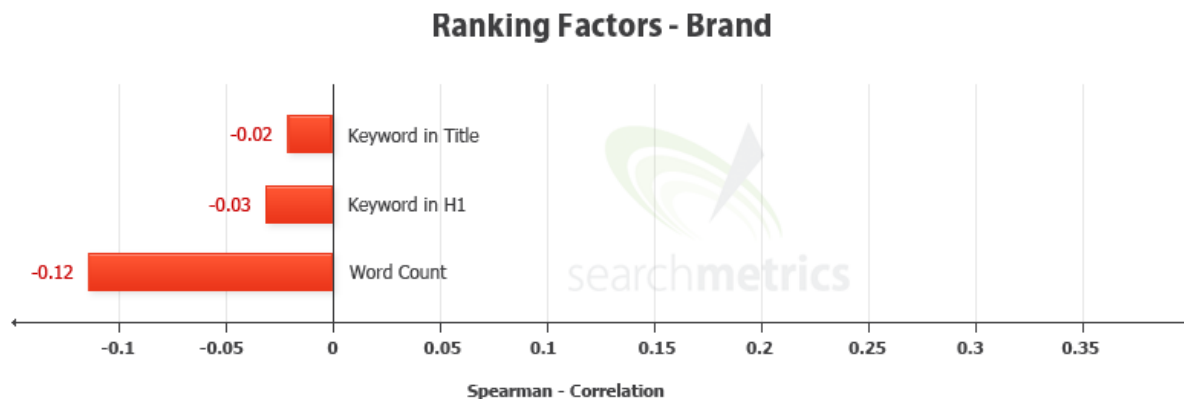
However, the point at which these factors will overcome the sheer quantity of links is still a long way off for many keywords. Taking the number of links depending on the ranking of the linked site from our monitored keywords you get the following values:



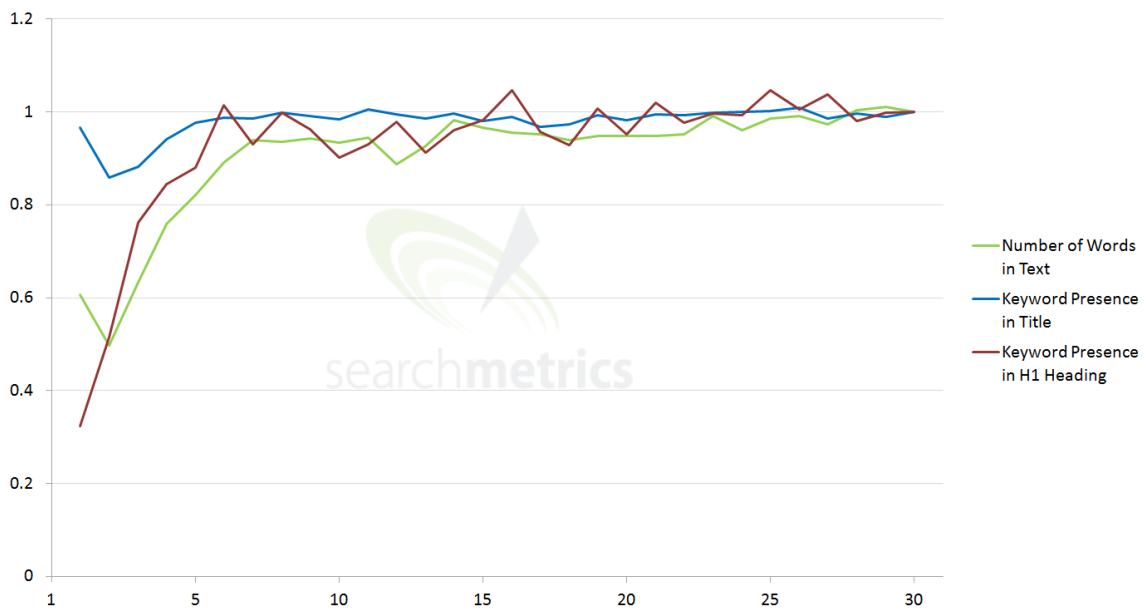
4. Brand power endures

For a while now, the rule among SEOs has been that brands enjoy a ranking advantage and that its particularly worthwhile if you can establish yourself as this type of brand. However, the 'brand' factor is difficult to establish in large-scale data analyses. In the end, it is nearly impossible to ascertain the thematic criteria for a brand without access to a search engine's algorithm. However, it's not entirely impossible.

For a few of our analyzed on-page criteria, the effect of brand-power is obvious and even seems to turn the 'conventional' SEO logic on its head. This is noticeable with the following elements, all of which (surprisingly) feature a negative correlation:



The core message of this graph is that the less often a keyword appears in the headline or title and the fewer the words in the text, the better a page will rank. In addition, text quantity seems to have no real positive effect. This is quite surprising at first. However, if you look at the precise trend of these metrics for the top 30 places a pattern emerges: right up to the top 10 results, the factors behave as expected without any large correlation in any direction. However, in first place, the natural niche for brands, everything is turned upside down:



To make things clearer: in this graph you can see an index value for the frequency of the assessed ranking factors on the y-axis. Google rankings appear on the x-axis. The pages with the lowest figures appear in the top three positions.

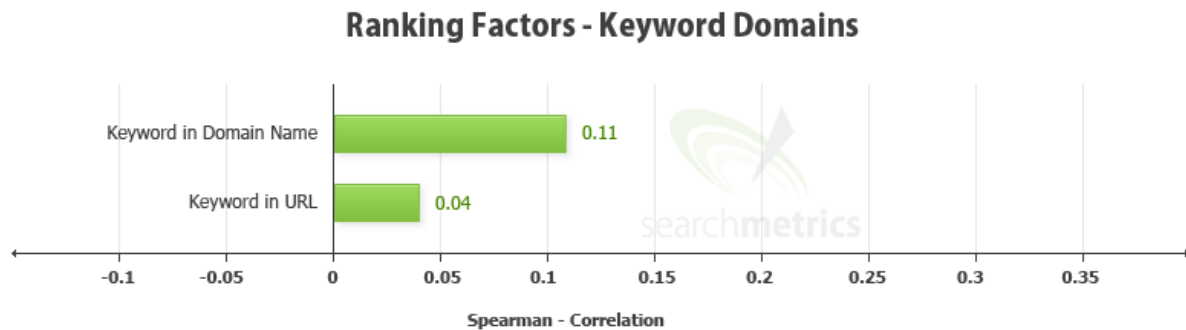
From this we can take it that to end up in the top positions you do not need a keyword in the title or in the H1 headline and even the average word count tends to drop off towards the top results – with the clear exception of the outlier at number one.

In this case, the data contradicts traditional on-page SEO theory. For example, we would normally expect that a title keyword represents a large advantage and in the past we have often assumed that keywords in the H1 tag had a positive influence. However, apparently these factors are now weighted completely differently in Google results. Strong brands rank in the top five even without perfectly conforming on-page structures.

This effect occurs throughout the whole of the analyzed keyword set and not just for brand and brand combination keywords.

5. Keyword domains & URL keywords

The power of keyword domains has been known for years and is still clearly visible in our analysis:

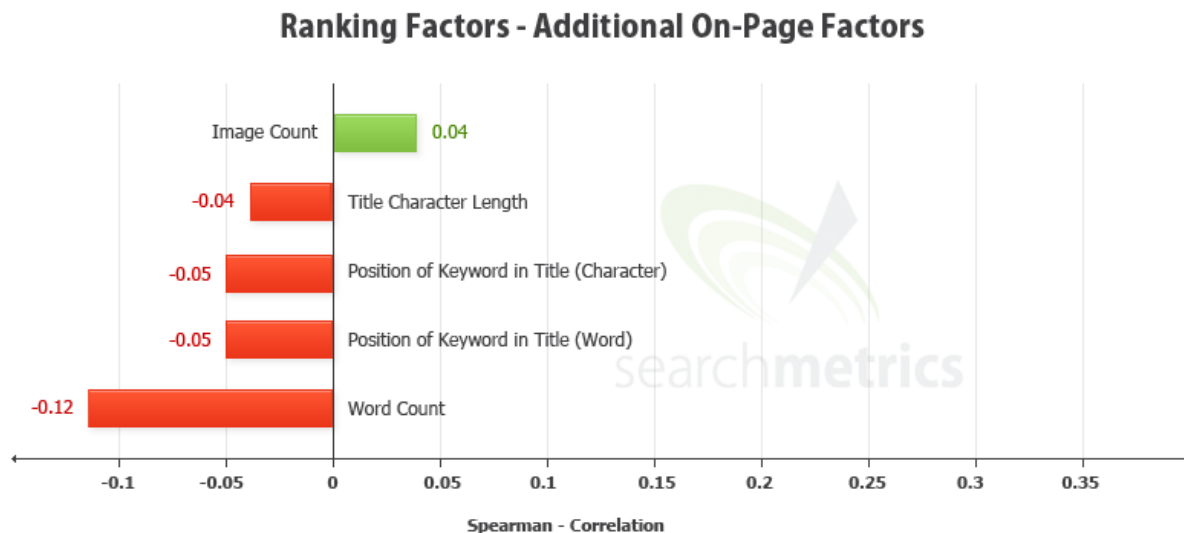


Keyword domains correlate much better with high rankings than results from just any random start page. The correlation with keyword domains is also significantly higher than the correlations for keywords in the rest of the URL. Although Google has repeatedly emphasized that these sites will slowly weaken in power, this does not yet seem to be the case.

6. Additional on-page factors

We have tried to examine as many factors as possible for their potential impact, at least as far as is possible with such a large sample base. Besides the factors than have been summarized in the main points above, there are still a number of other identified factors that differ little from the expected results or not at all.

Of these the following on-page factors can be included:



Although you might think that pages with a lot of multimedia content would tend to rank better (possibly also indirectly through better user signals e.g. social media links), there was no strong correlation in our analysis between rankings and sites with more images (where all images excluding 'spacers' were counted).

Just as unexpected is the negative correlation between rankings and text length. In this case, the correlation between brands and higher rankings contradicts the SEO logic presented above.

Title length and keyword title position (according to character and word) featured slightly negative correlations. This corresponds with the experience that keywords placed earlier tend to be more strongly weighted and individual keywords have less weight in longer titles.

Be careful about drawing conclusions: correlation ≠ causation!

We would like to emphasize that this does not in any way guarantee that corresponding factors have an effect on rankings or that they are even used by Google as signals. Questions like "does a site receive social signals because it ranks well or does it rank well because it receives social signals?" are absolutely valid and cannot be answered unequivocally with the current data.

Some information regarding our data

For our dataset we selected an extremely large keyword set of 10,000 search terms from Google.com U.S. However, we did not just include the top 10,000 search terms according to search volume, since they contain a disproportionately high number of brand keywords which might have distorted the assessment of many other key factors. Instead our reference dataset includes:

- A mix of different keywords with, if not the largest, then at least generally high search volumes,
- around 1 in 10 are keywords that were identified as navigation-oriented according to our logic,
- the rest are a mix of keywords from a variety of CPC areas to best cover transactional (higher CPC) and information oriented (lower CPC) searches as well as the hybrids in between.

The assessment was limited to organic searches – AdWords, Universal Search OneBoxes, 2 to 7 packs, sitelinks, iGoogle integrations etc. were not included.

The analysis' 10,000 analyzed keywords lead to:

- 30,000 SERPs
- with 300,000 titles, descriptions and URLs.

The ranking sites' content included:

- 14.68 GB in data,
- 92,672 AdSense blocks,
- 338,562,612 Facebook comments, 3.04 billion shares and 8.1 billion likes