The Theory Behind Ranking Factors – Whiteboard Friday

Since day one of SEO, marketers have tried to determine what factors Google takes into account when ranking results on the SERPs. In this brand new Whiteboard Friday, Russ Jones discusses the theory behind those ranking factors, and gives us some improved definitions and vocabulary to use when discussing them.

Click on the whiteboard image above to open a high resolution version in a new tab!

Video Transcription

Hi, folks. Welcome back to another Whiteboard Friday. Today, we're going to be talking about ranking factors and the theory behind them, and hopefully get past some of these – let's say controversies – that have come up over the years, when we've really just been talking past one another.

You see, ranking factors have been with us since pretty much day one of search engine optimization. We have been trying as SEOs to identify exactly what influences the algorithm. Well, that's what we're going to go over today, but we're going to try and tease out some better definitions and vocabulary so that we're not talking past one another, and we're not constantly beating each other over the heads about correlation and not causation, or some other kind of nuance that really doesn't matter.

Direct

So let’s begin at the beginning with direct ranking factors. This is the most narrow kind of understanding of ranking factors. It’s not to say that it’s wrong – it’s just pretty restrictive. A direct ranking factor would be something that Google measures and directly influences the performance of the search result.

So a classic example would actually be your robots.txt file. If you make a change to your robots.txt file, and let’s say you disallow Google, you will have a direct impact on your performance in Google. Namely, your site is going to disappear.

The same is true for the most part with relevancy. Now, we might not know exactly what it is that Google is using to measure relevancy, but we do know that if you improve the relevancy of your content, you’re more likely to rank higher. So these are what we would call direct ranking factors. But there’s obviously a lot more to it than that.

Google has added more and more features to their search engine. They have changed the way that their algorithm has worked. They’ve added more and more machine learning. So I’ve done my best to try and tease out some new vocabulary that we might be able to use to describe the different types of ranking factors that we often discuss in our various communities or online.

Indirect

Now, obviously, if there are direct ranking factors, it seems like there should be indirect ranking factors. And these are just once-removed ranking factors or interventions that you could take that don’t directly influence the algorithm, but they do influence some of the direct ranking factors which

influence the algorithm.

I think a classic example of this is hosting. Let's say you have a site that's starting to become more popular and it's time to move off of that dollar-a-month cPanel hosting that you signed up for when you first started your blog. Well, you might choose to move to, let's say, a dedicated host that has a lot more RAM and CPU and can handle more threads so everything is moving faster.

Time to first byte is faster. Well, Google doesn't have an algorithm that's going out and digging into your server and identifying exactly how many CPU cores there are. But there are a number of direct ranking factors, those that are related perhaps to user experience or perhaps to page speed, that might be influenced by your hosting environment.

Subsequently, we have good reason to believe that improving your hosting environment could have a positive influence on your search rankings. But it wouldn't be a direct influence. It would be indirect.

The same would be true with social media. While we're pretty sure that Google isn't just going out and saying, "Okay, whoever is the most popular on Twitter is going to rank," there is good reason to believe that investing your time and your money and your energy in promoting your content on social media can actually influence your search results.

A perfect example of this would be promoting an article on Facebook, which later gets picked up by some online publication and then links back to your site. So while the social media activity itself did not directly influence your search results, it did influence the links, and those links influenced your search results.

So we can call these indirect ranking factors. For politeness' sake, please, when someone talks about social media as a ranking factor, just don't immediately assume that they mean

that it is a direct ranking factor. They very well may mean that it is indirect, and you can ask them to clarify: "Well, what do you mean? Do you think Google measures social media activity, or are you saying that doing a better job on social is likely to influence search results in some way or another?"

So this is part of the process of teasing out the differences between ranking factors. It gives us the ability to communicate about them in a way in which we're not, let's say, confusing what we mean by the words.

Emergent

Now, the third type is probably the one that's going to be most controversial, and I'm actually okay with that. I would love to talk in either the comments or on Twitter about exactly what I mean by emergent ranking factors. I think it's important that we get this one clear in some way, shape, or form because I think it's going to be more and more and more important as machine learning itself becomes more and more and more important as a part of Google's algorithm.

Many, many years ago, search engine optimizers like myself noticed that web pages on domains that had strong link authority seemed to do well in organic search results, even when the page itself wasn't particularly good, didn't have particularly good external links – or any at all, and even didn't have particularly good internal links.

That is to say it was a nearly orphaned page. So SEOs started to wonder whether or not there was some sort of domain-level attribute that Google was using as a ranking factor. We can't know that. Well, we can ask Google, but we can only hope that they'll tell us.

So at Moz, what we decided to do was try and identify a series of domain-level link metrics that actually predict the

likelihood that a page will perform well in the search results. We call this an emergent ranking factor, or at least I call it an emergent ranking factor, because it is obviously the case that Google does not have a specific domain-authority-like feature inside their algorithm.

But on the contrary, they also do have a lot of data about links pointing to different pages on that same domain. What I believe is going on is what I would call an emergent ranking factor, which is where, let's say, the influence of several different metrics – none of which have a particularly intended purpose of creating something – end up being easy to measure and to talk about as an emergent ranking factor, rather than as part of all of its constituent elements.

Now, that was kind of a mouthful, so let me give you an example. When you're making a sauce if you're cooking, one of the most common parts of that would be the production of a roux. A roux would be a mix, normally of equal weights of flour and fat, and you would use this to thicken the sauce.

Now, I could write an entire recipe book about sauces and never use the word "roux". Just don't use it, and describe the process of producing a roux a hundred times, but never actually use the word "roux", because "roux" describes this intermediate state. But it becomes very, very useful as a chef to be able to just say to another chef (or a sous-chef, or a cook in their cookbook), "produce a roux out of" and then whatever is the particular fat that you're using, whether it's butter or oil or something of that sort.

So the analogy here is that there isn't really a thing called a roux that's inside the sauce. What's in the sauce is the fat and the flour. But at the same time, it's really convenient to refer to it as a roux. In fact, we can use the word "roux" to know a lot about a particular dish without ever talking about the actual ingredients of flour and of fat.

For example, we can be pretty confident that if a roux is called for in a particular dish, that dish is likely not bacon because it's not a sauce. So I guess what I'm trying to get at here is that a lot of what we're talking about with ranking factors is using language that is convenient and valuable for certain purposes.

Like DA is valuable for helping predict search results, but it doesn't actually have to be a part of the algorithm in order to do that. In fact, I think there's a really interesting example that's going on right now – and we're about to see a shift from the categories – which are Core Web Vitals.

Google has been pushing page speed for quite some time and has provided us several iterations of different types of metrics for determining how fast a page loads. However, what appears to be the case is that Google has decided not to promote individual, particular steps that a website could take in order to speed up, but instead wants you to maximize or minimize a particular emergent value that comes from the amalgamation of all of those steps.

We know that the three different types of Core Web Vitals are: first input delay, largest contentful paint, and cumulative layout shift. So let's talk about the third one. If you've ever been on your cell phone and you've noticed that the text loads before certain other aspects and you start reading it and you try and scroll down and as soon as put your finger there an ad pops up because the ad took longer to load and it's just jostling the page, well, that's layout shift, and Google has learned that users just don't like it. So, even though they don't know all of the individual factors underneath that are responsible for cumulative layout shift, they know that there's this measurement, that explains all of it, that is great shorthand, and a really effective way of determining whether or not a user is going to enjoy their experience on that page.

This would be an emergent ranking factor. Now, what's interesting is that Google has now decided that this emergent ranking factor is going to become a direct ranking factor in 2021. They're going to move these descriptive factors that are amalgamations of lots of little things and make them directly influence the search results.

So we can see how these different types of ranking factors can move back and forth from categories. Back to the question of domain authority. Now, Google has made it clear they don't use Moz's domain authority – of course they don't – and they do not have a domain-authority-like metric. However, there's nothing to say that at some point they could not build exactly that, some sort of domain-level, link-based metric which is used to inform how to rank certain pages.

So an emergent ranking factor isn't stuck in that category. It can change. Well, that's enough about emergent ranking factors. Hopefully, we can talk more about that in the comments.

Validating

The next type I wanted to run through is what I would call a validating ranking factor. This is another one that's been pretty controversial, which is the Quality Rating Guidelines' list of things that matter, and probably the one that gets the most talked about is E-A-T: Expertise, Authority, and Trustworthiness.

Well, Google has made it clear that not only do they not measure E-A-T (or at least, as best as I've understood, they don't have metrics that are specifically targeted at E-A-T), not only do they not do that, they also, when they collect the data from quality raters on whether or not the SERPs they're looking at meet these qualifications, they don't train their algorithm against the labeled data that comes back from their quality raters, which, to me, is surprising.

It seems to me like if you had a lot of labeled data about quality, expertise, and authoritativeness, you might want it trained against that, but maybe Google found out that it wasn't very productive. Nevertheless, we know that Google cares about E-A-T, and we also have anecdotal evidence.

That is to say webmasters have noticed over time, especially in "your money or your life" types of industries, that expertise and authority does appear to matter in some way, shape, or form. So I like to call these validating ranking factors because Google uses them to validate the quality of the SERPs and the sites that are ranking, but doesn't actually use them in any kind of direct or indirect way to influence the search results.

Now, I've got an interesting one here, which is what I would call user engagement, and the reason why I've put it here is because this still remains to be a fairly controversial ranking factor. We're not quite sure exactly how Google uses it, although we do get some hints every now and then like Core Web Vitals.

If that data is collected from actual user behavior in Chrome, then we've got an idea of exactly how user engagement could have an indirect impact on the algorithm because user engagement measures the Core Web Vitals, which, coming in 2021, are going to directly influence the search results.

Correlation

So validating is this fourth category of ranking factors, and the last – the one that I think is the most controversial – is correlates. We get into this argument every time: "correlation does not equal causation", and it seems to me to be the statement that the person who only knows one thing about statistics knows, and so they always say it whenever anything ever comes up about correlation.

Yes, correlation does not imply causation, but that doesn't mean it isn't very, very useful. So let's talk about social metrics. This is one of the classic ones. Several times we've run various studies of ranking factors and discovered a direct relationship – a strong relationship – between things like Facebook likes or Google pluses in rankings.

All right. Now, pretty much everyone immediately understood that the reason why a site would have more plus-ones in Google+ and would have more likes in Facebook would be because they rank. That is to say, it's not Google going out and depending on Facebook's API to determine how they're going to rank the sites in their search engine.

On the contrary, performing well in their search engine drives traffic, and that traffic then tends to like the page. So I understand the frustration there when customers start asking, "Well, these two things correlate. Why aren't you getting me more likes?"

I get that, but it doesn't mean that it isn't useful in other ways. So I'll give you a good example. If you are ranking well for a keyword but yet your social media metrics are poorer than your competitors', well, it means that there's something going on in that situation that is making your users engage better with your competitors' sites than your own, and that's important to know.

It might not change your rankings, but it might change your conversion rate. It might increase the likelihood that you get found on social media. Even more so, it could actually influence your search results. Because, when you recognize the reason why you're not getting any likes to your page is because you have broken code, so the Facebook button isn't working, and then you add it and you start getting shared and more and more people are engaging with and linking to your content, well, then we start having that indirect effect on your rankings.

So, yeah, correlation isn't the same as causation, but there's a lot of value there. There's a new area that I think is going to be really, really important for this. This is going to be natural language processing metrics. These are various different technologies that are on the cutting edge. Well, some are older. Some are newer. But they allow us to kind of predict how good content is.

Now, chances are we are not going to guess the exact way that Google is measuring content quality. I mean, unless a leaked document or something shows up, we're probably not going to get that lucky. But that doesn't mean we can't be really productive if we have a number of correlates, and those correlates can then be used to guide us.

So I drew a little map here to kind of serve as an example. Imagine that it's the evening and you're camping, and you decide to go on a quick hike, and you take with you, let's say, a flag or a series of flags, and you mark the trail as you go so that when it gets later, you can flick on your flashlight and just follow the flags, picking them up, to lead you back to camp.

But it gets super dark, and then you realize you left your flashlight back at camp. What are you going to do? Well, we need to find a way to guide ourselves back to camp. Now, obviously, the flags would have been the best situation, but there are lots of things that are not the camp itself and are not the path itself, but would still be really helpful in getting us back to camp. For example, let's say that you had just put out the fire after you left camp. Well, the smell of the smoke is a great way for you to find your way back to the camp, but the smoke isn't the camp. It didn't cause the camp. It didn't build the camp. It's not the path. It didn't create the path. In fact, the trail of smoke itself is probably quite off the path, but once you do find where it crosses you, you can follow that scent. Well, in that case, it's really valuable even though it just mildly correlates with exactly

where you need to get.

Well, the same thing is true when we're talking about something like NLP metrics or social media metrics. While they might not matter in terms of influencing the search results directly, they can guide your way. They can help you make better decisions. The thing you want to stay away from is manipulating these types of metrics for their own sake, because we know that correlates are the furthest away from direct ranking factors – at least when we know that the correlate itself is not a direct ranking factor.

All right. I know that's a lot to stomach, a lot to take in. So hopefully, we have some material for us to discuss below in the comments, and I look forward to talking with you more. Good luck. Bye.

Video transcription by Speechpad.com

https://feedpress.me/link/9375/13886331/ranking-factors-theory