

Search engines think I plagiarized my own content! My Hacker News Case Study (UPDATE V4)

Search engines think I plagiarized my own content. Being the author, the plagiarist and the victim is a strange experience, especially when you are a senior organic growth strategist!

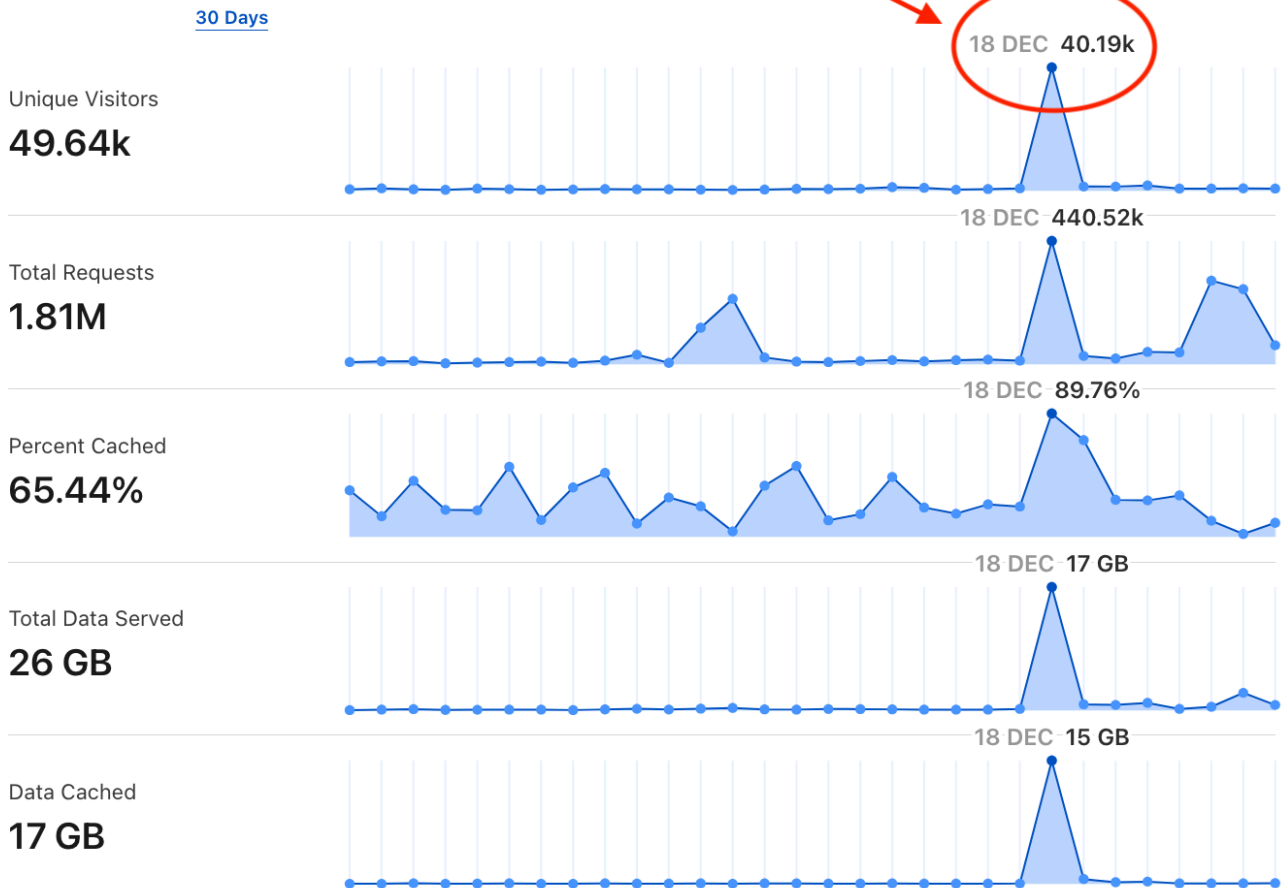
It all started on December 18, 2024 after I shared one of my articles entitled [Google, the search engine that's forgotten how to search](#) on Hacker News. This was my first time on HN. My article got 100+ comments in a few hours and the upvotes reached 80 points.

Sure, that's not much but I was happy because I discovered an intellectually curious community, clearly more technically proficient than most! And then, my critical article about Google Search got shadow-banned by... Google. Yes, many use this expression carelessly. I don't. I'll focus on Google because it is the leading search engine in terms of market share but this goes beyond Google. Together, we'll try to understand what happened.

Along the way, I'll share contrarian and hopefully insightful observations. I wrote some takeaways for you, your blog, your brand at the end. I hope you will disagree with me and explain why I'm wrong because I believe friction and open debate are among the fastest ways to grow.

Here's my Cloudflare web traffic graph—which counts all types of traffic, including bots/spiders/crawlers. The surge in referral traffic occurred minutes after publishing on Hack News.

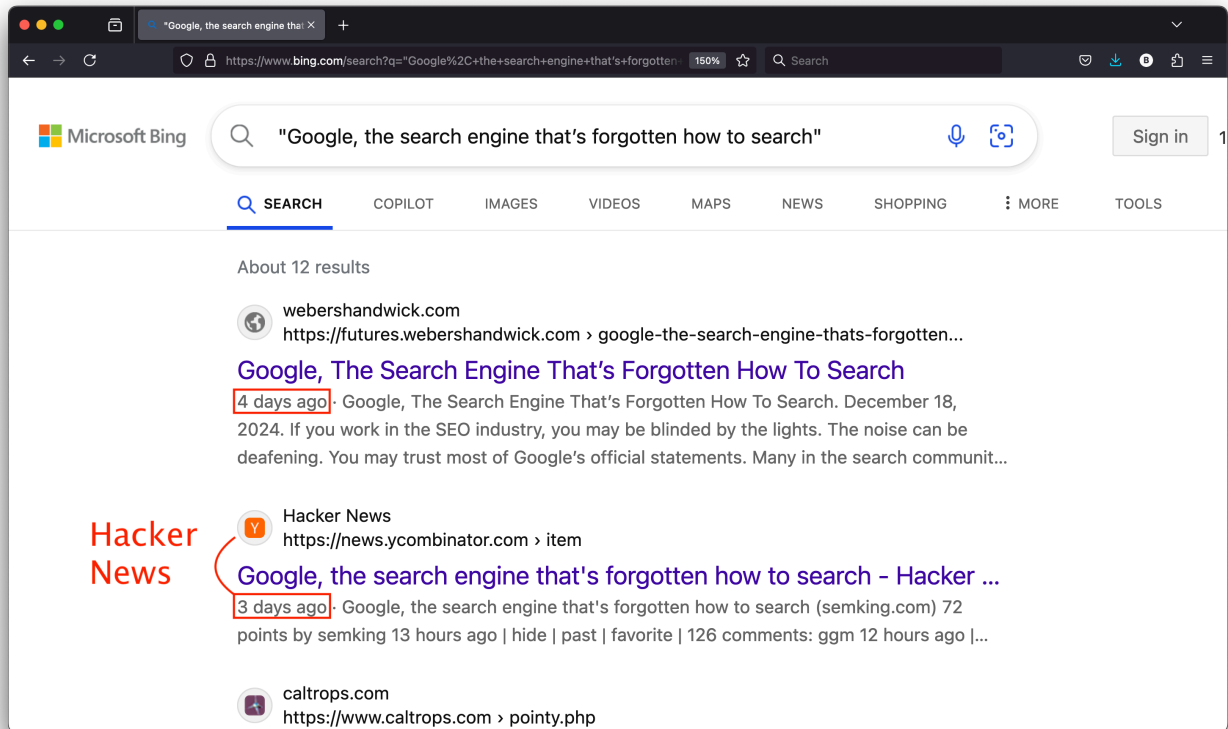
Article published on
December 18, 2024



I could only confirm about 2,200 human readers on December 18, 2024 —about 5% of the 40.19k “unique visitors” shown in my Cloudflare graph. Yes, I have a strict set of criteria. But I was satisfied to see thousands read my article!

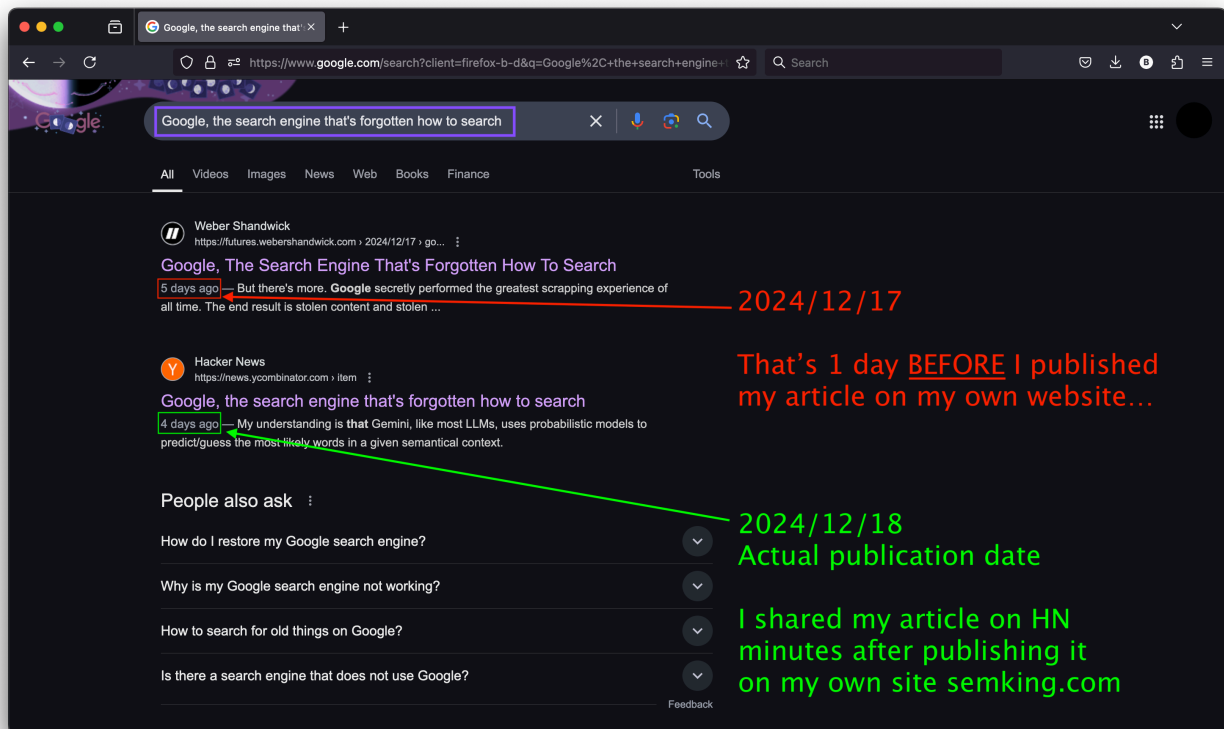
I’m used to the lovely and supportive LinkedIn audience that usually claps without reading. You didn’t clap but you read everything and that was heartwarming. I know because I was tracking different KPIs behind the scenes: heatmaps, session times, people translating my content, etc. While monitoring the many on-page and off-page metrics, I uncovered a potentially parasitic ecosystem. It gravitates around HN and it is anything but organic.

Days after sharing my URL on Hacker News, the futures[.]webershandwick[.]com subdomain climbed at the top of Bing’s organic search results for my exact title. And my content appeared a day BEFORE I finished writing my article. I’m often surprised by my own abilities but being published before writing my content didn’t make any sense! 🤔



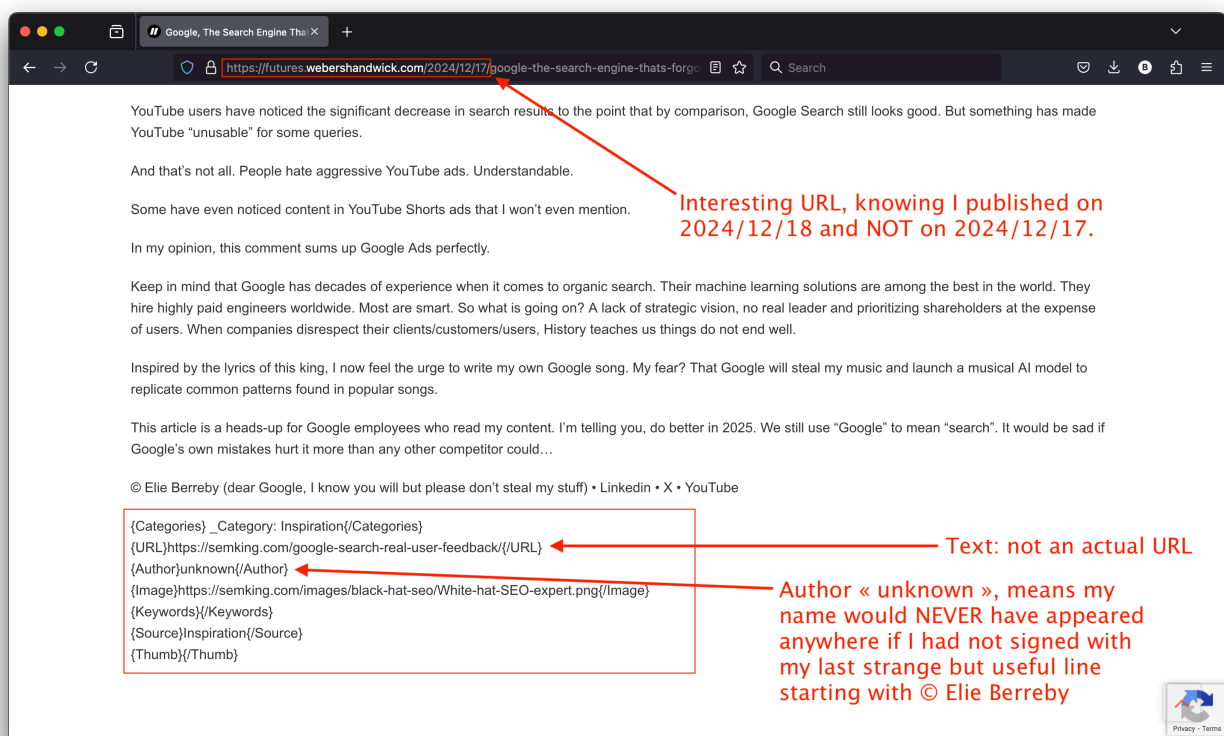
My first reaction? The Weber Shandwick brand probably used their “futures” subdomain for syndication and for some strange reason, the publication date in Bing Search was one day before my publication date. It must have been a Bing mistake. Surely, Google Search would never fall for this.

I checked the next day and my article ranked FIRST on Google Search but NOT on my website: on this third-party subdomain. And once again, with a publication date of December 17, 2024 —one day before I finished writing it on December 18, 2024!



What led to this unfortunate turn of events? I remembered the old saying and I didn't want to attribute to malice what could be explained by stupidity. That's not me calling anyone stupid (yet). That's just the saying.

Here, we have a brand trusted by search engines that leverages its Site Reputation, programmatically detects popular Hacker News content and injects it into its website hosted on its "futures" subdomain...



Look at my scraped article on their website. You'll notice the bottom text is formatted in a way that suggests it's part of an HTML or XML structure. Nothing directly indicates a failure in API calls but let's read everything:

- **{Categories}_Category: Inspiration{/Categories}**: categorizes my content under "Inspiration", a category that does not exist on my website. Yes, I felt inspired... but this was created by WS for their site. Glad to see my content was inspiring 😊
- **{URL}https://semking[.]com/google-search-real-user-feedback/{/URL}**: this was supposed to be the URL linking to my article but it is just plain text. No backlink! 😞
- **{Author}unknown{/Author}**: this indicates that the author of the content, me, is unknown. Interesting! 😏
- **{Keywords}/{/Keywords}**: this "tag" is empty, no keywords have not been specified or retrieved.
- **{Source}Inspiration{/Source}**: apparently, the source of the content was "Inspiration"... Inspiration, again? The real source was my inspired little brain and the hours I spent researching and writing my previous article...
- **{Thumb}/{/Thumb}**: this "tag" is also empty, probably because they tried to generate a thumbnail image of my article and failed. I'm giving them a well-deserved thumb down! 😞

I guess their API or the system generating this output is misconfigured. Errors in the parsing logic can lead to such incomplete or malformed output. But what matters isn't the cause but the real-world consequences: no author (unknown), no attribution, and no backlink to the original source (my URL).

And then, there's the date issue. Why December 17 when I originally published on December 18? We all know how easy it is to tweak the publication date of our articles, right? But which respectable brand would intentionally manipulate Search Engine Result Pages (SERPs) by scraping content and setting the publication date one day BEFORE the actual date?

I'm sure Weber Shandwick didn't change the date, even programmatically. So, why was the date of publication on their website set to the previous day? Even their URL includes 2024/12/17 —and I published on December 18, 2024!

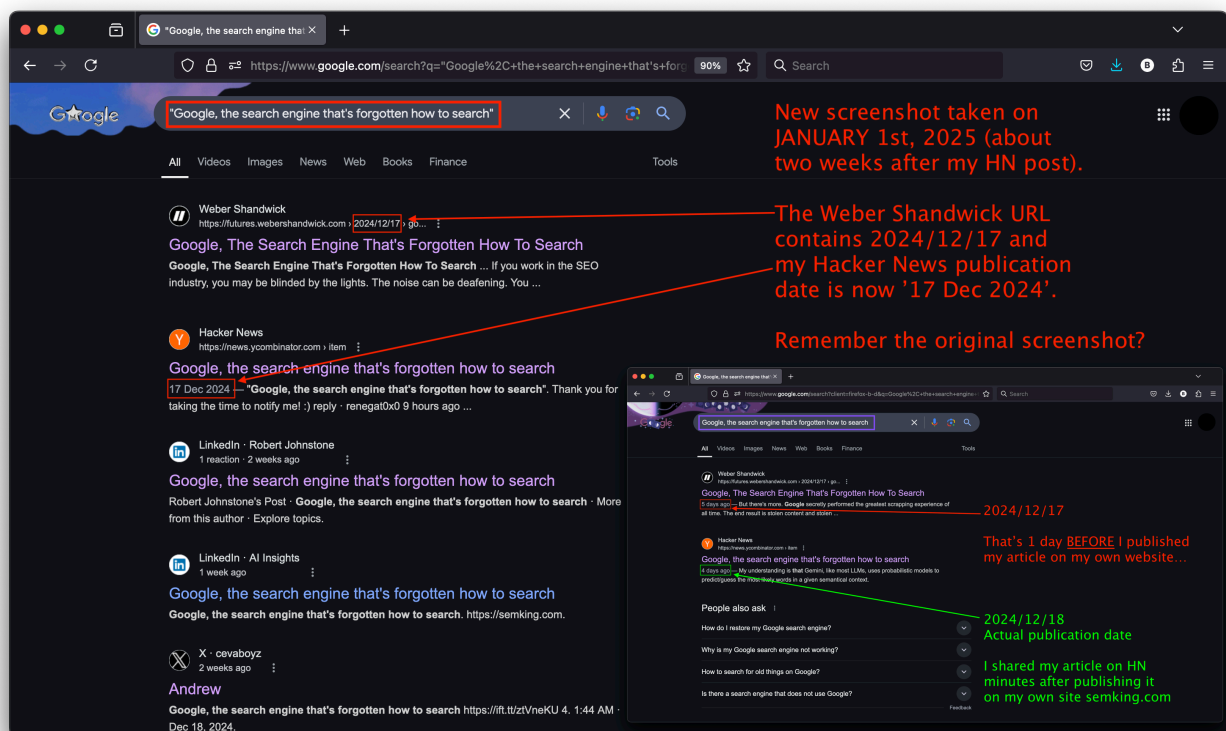
This does look intentional. But is the probability that this was unintentional above zero? Actually, the probability this was unintentional is close to 1. I'm the "victim" here but I'll take the opposite side and "defend" the Weber Shandwick brand. That's always worth doing because we should all avoid arriving at erroneous conclusions.

The superficial answer? The time zone difference probably caused this issue. You see, I live in Europe but I've been working in Eastern Time for the past few years. That means we have different local times: if I publish an article at 2 AM (local time in the morning for me), it is still the previous day for people in the United States. And I published my last article in the early morning hours (European time). Great! Problem solved, right?

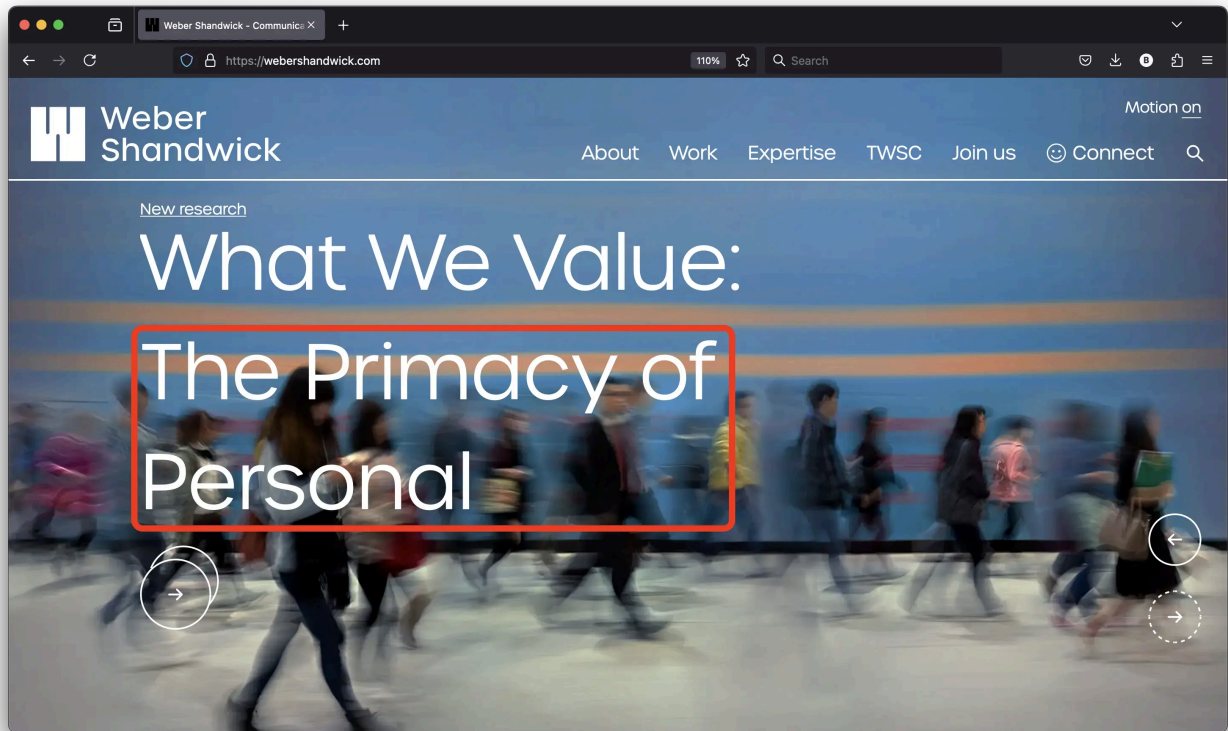
Sadly, not at all! This does NOT explain why Google Search perceived the WS website as the first source to publish my content. Remember: my content was first published on my website, then on Hacker News and then scraped and republished on the third-party WS subdomain.

I took a new screenshot of Google Search's SERPs on January 1, 2025. Look at how the Google search engine now displays both WS and HN as "17 Dec 2024" entries! Initially, Google's SERPs showed a 1-day latency between the two entries. The chronology makes no sense and I'm afraid I can't think of a plausible explanation.

This looks like a search engine issue: compare it with my original screenshot of Google's SERPs.



While something seems wrong on the search engines' side, nothing justifies detecting popular content on Hacker News, instantly scraping the text and then failing to give attribution to authors. When this happens intentionally, it demonstrates malicious intent. But a series of failed programmatic processes could also lead to the same result. And the original author will be penalized. Copying content without attribution is not acceptable. This hurts tiny content creators.



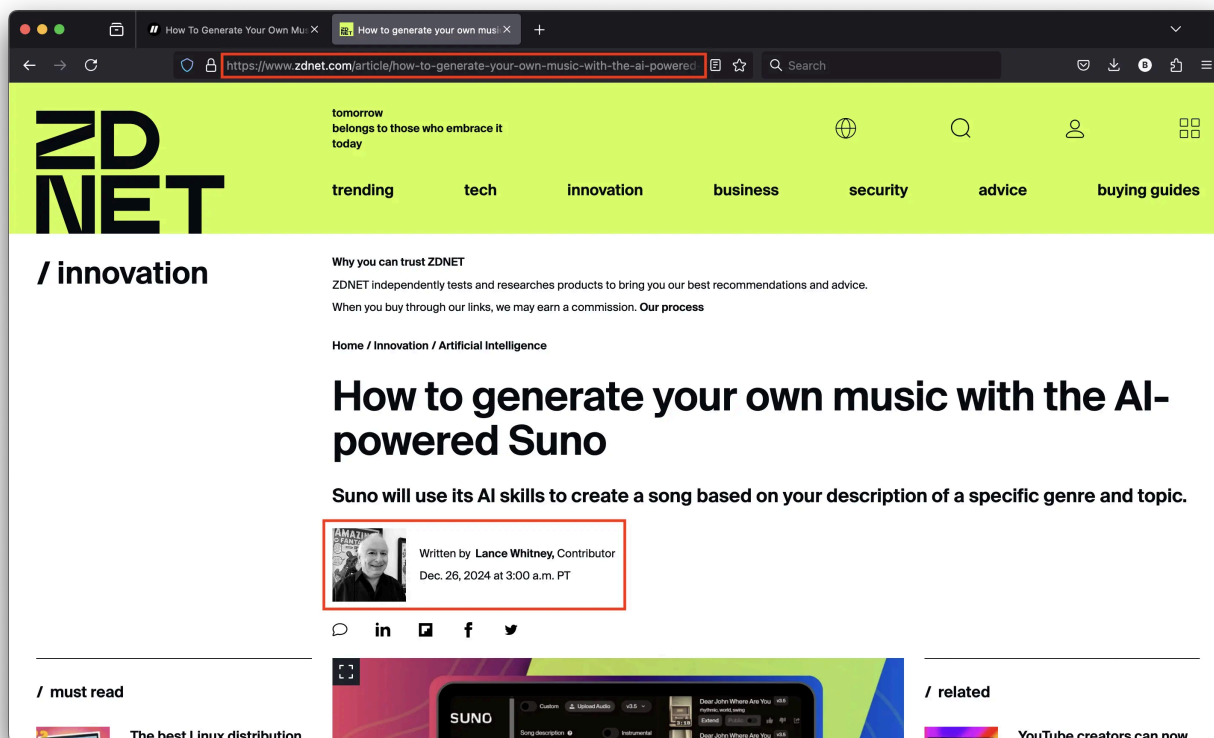
When done properly, syndication can amplify the original author's reach and visibility. Done poorly, it can ruin everything. Please fix this and value the primacy of personal... blogs! 🙄

From a search engine's perspective, the Weber Shandwick (WB) subdomain is a website that has a higher trust factor than small personal websites. What is the risk if the WB system automatically detects, scrapes and instantly re-publishes your popular HN content without giving you attribution?

Simply put, your original content might end up being crawled by search engines on your own website AFTER being first crawled on the THIRD-PARTY website. To search engines, you may therefore appear to be the one plagiarizing your own content! Pretty ironic, right?

Usually, "content cannibalization" means your website contains multiple pieces of content that are all showing up for similar search intents. Your content eats itself, but on your own website: one of your secondary pages may appear more often than the intended page. But it is still YOU, YOUR content, YOUR website. Now, if you are a tiny brand wrongfully "convicted" of plagiarism, other sources can cannibalize your organic visibility using YOUR content!

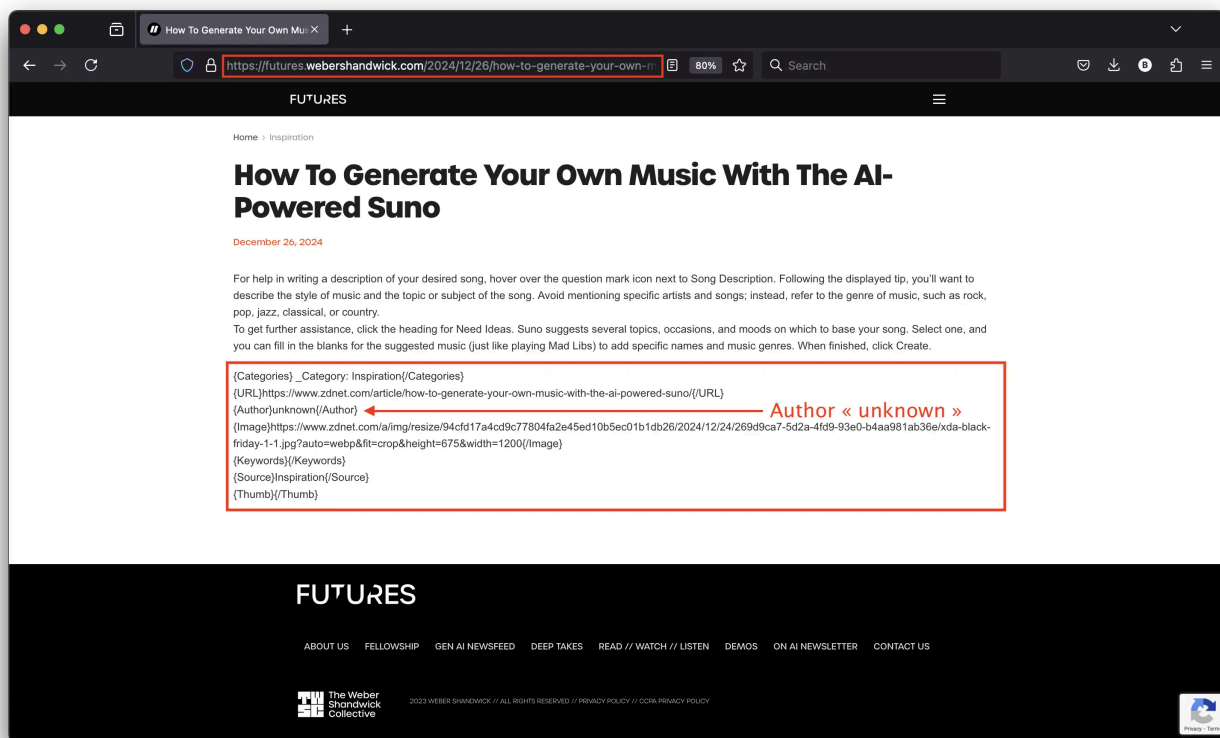
I'll now explain the importance of online reputation and crawling velocity. Suppose you are a large established brand publishing often —such as ZDNET. Imagine one of your contributors named "Lance" wrote an article about an AI model to artificially "generate" music. Please, don't get me started on this terrible topic! 🙄



As a ZDNET author, would Lance suffer as much as a tiny content creator if a third party attempted to position themselves as quickly as possible after the original publication? No, because popular publishers such as ZDNET have large websites. And because they publish often, they benefit from a high crawl budget.

That means search engine spiders such as Googlebot crawl and re-crawl very often. The crawler's objective is to detect changes to existing pages, to discover new content and new links. Any author, even unknown, who would publish new content on a trusted website such as ZDNET would get their article crawled and indexed by search engines almost immediately.

Lance's article also got scrapped and it was immediately republished on the futures[.]webershandwick[.]com subdomain.



Lance, who was very clearly a ZDNET contributor on the original website, is now an “unknown” author on the “futures” subdomain of Weber Shandwick. What about his ZDNET article’s URL? Plain text again, not a URL. Same issues. But at least the date is correct this time, probably because both ZDNET and WS use U.S. time zones.

The good news? The exact same type of behavior (scraping/republishing/no attribution) will have different effects. The same actions might not lead to the same consequences, depending on the size, online reputation and authority of the source being “syndicated” (I meant “scrapped” but whatever). Scraping ZDNET will have almost no consequence: they are too BIG to suffer.

Site quality score: the secret metric

Search engines secretly attribute a website quality score to your web property. Many internal and external factors influence this ever-evolving score. If your score is extremely low, you’ll have to work hard to be “re-evaluated” and it will take time to raise the bar. If your score is incredibly high, you’d really have to intentionally mess up pretty badly over an extended period of time to suffer from negative results.

Suppose you went to prison. Once out, you’d have to prove yourself to everyone who knows your past. People would be suspicious. Nobody would instantly trust you. And even if you managed to earn some level of trust, your past would remain in the back of our collective mind. That’s what publishers with a low site quality score deal with. That’s why many would instead start from scratch and build their reputation using another domain. Some call this “rebranding” 😊

But the comparison stops here. In the real world you can spend a lifetime building a reputation and destroy everything in 5 minutes. Online things are different. Suppose you see the worst possible content in a LinkedIn post tomorrow. The person would probably get banned from the social platform but would the site quality score of the entire LinkedIn[.]com website go down? No. What if a seller managed to list something illegal on Amazon[.]com? Would it impact the site quality score? No, it might have an effect at a page level but it would not damage the entire site score.

The site quality score is one of the most unknown and yet one of the most important topics. Regarding the timeline, the sooner you demonstrate credibility and trustworthiness, the better.

Your content VS our search intent

Want good news? Your content is safe if it is generic, bland, fluffy and uninteresting. Trust me, even bots won't want to plagiarize an article about the 7 secrets marrying your cat taught you about B2B sales! 😊

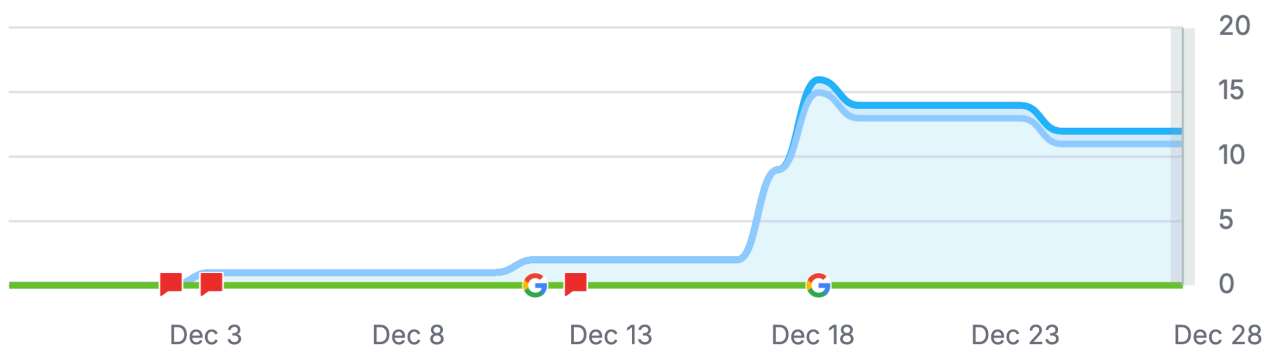
I couldn't care less about the traditional SEO key performance indicators. I didn't perform any keyword research before writing my previous article because I knew it could influence my choice of words. I didn't even use an h1 HTML tag initially. I didn't opt for an SEO optimized title. I don't care about links unless they are qualitative.

Of course, not doing any keyword research was a mistake because it means I'll rank for unexpected keywords and random search queries. But why is ranking for random search queries risky? Because your organic traffic won't be targeted.

When an individual searches, they always have an intention. Sometimes they aren't aware of their own intention but it is there. And our intentions quickly evolve over time and searches. While search engines might initially give your content a rank boost because it is fresh, they'll keep analyzing the users' search intent and if they notice people are dissatisfied, you'll progressively lose organic positions. It actually makes sense: why rank content that does not match the users' search intent?

Organic Keywords 12

✓ Top 3 ✓ 4-10 ✓ 11-20 ✓ 21-50 ✓ 51-100 ✓ SERP Features



Do I care? No. I create content for human beings, not for bots. And like you, I hate search engine spam. Of course, when I advise clients, I'm paid to think and care about these things. I cannot reverse-engineer the predictive analytics models used by search engines to predict how content is expected to perform but I can create my own models. I can also try to anticipate the positive or negative impact of a deviation from the expected baseline CTR. Not an easy task in a zero-click world.

Inorganic & manipulated SERPs

I spent days working on this case study as a public service. I want to spark a debate about how fake the Search industry has become. I'm using my own example to highlight one of the many issues that affect almost all tiny content creators who are disappearing from organic results at a record pace. This might seem self-centered, but I couldn't use any other example.

Most brands would NEVER share their traffic and graphs publicly. Understandable: that's data that should never end up in your competitors' hands. How do I know? I constantly sign NDAs with global brands I advise. Sometimes, the data is sensitive.

And sometimes, the data is funny. I can't tell you how often I saw partial tracking. That means for years, people in large organizations have used and made assumptions based on incomplete tracking data. A very senior expert would see this in minutes but some still think you are an "expert" because you've spent 15 years doing something. Plot twist: that's not how things work. You are an expert when you are an outlier, when you think differently.

Back to me. This is my "personal brand". I despise the expression "personal branding" because we're people. But online, I'm my own brand. And I can share what I want to share—even and maybe especially when that does not make me look good!

I know how to rank anything at the top of search engines by using unethical tactics but I've always refused to do it. That never lasts and I hate shortcuts. Strategy always beats tactics.

I'm not virtue signaling, I'm just preparing you for what's coming. To write this "case study", I've opted for a passive approach.

At this early stage I'm passively monitoring and observing everything in SERPs without taking any impactful action. Again, this is a luxury I can afford when dealing with my personal site. A client would require confidentiality and immediate action. But I wanted to go through what most startups or personal blogs would go through when faced with this.

Very few have advanced SEO knowledge. Most self-proclaimed "SEO experts" are only experts in their own grandiose imagination. In practice, most have no clue. They don't understand search engines, vector embeddings or what behavioral patterns reveal from a psychological perspective. But they call themselves "experts" and by doing so, they damage the few who have real expertise (those who have put in the work when nobody was watching or clapping).

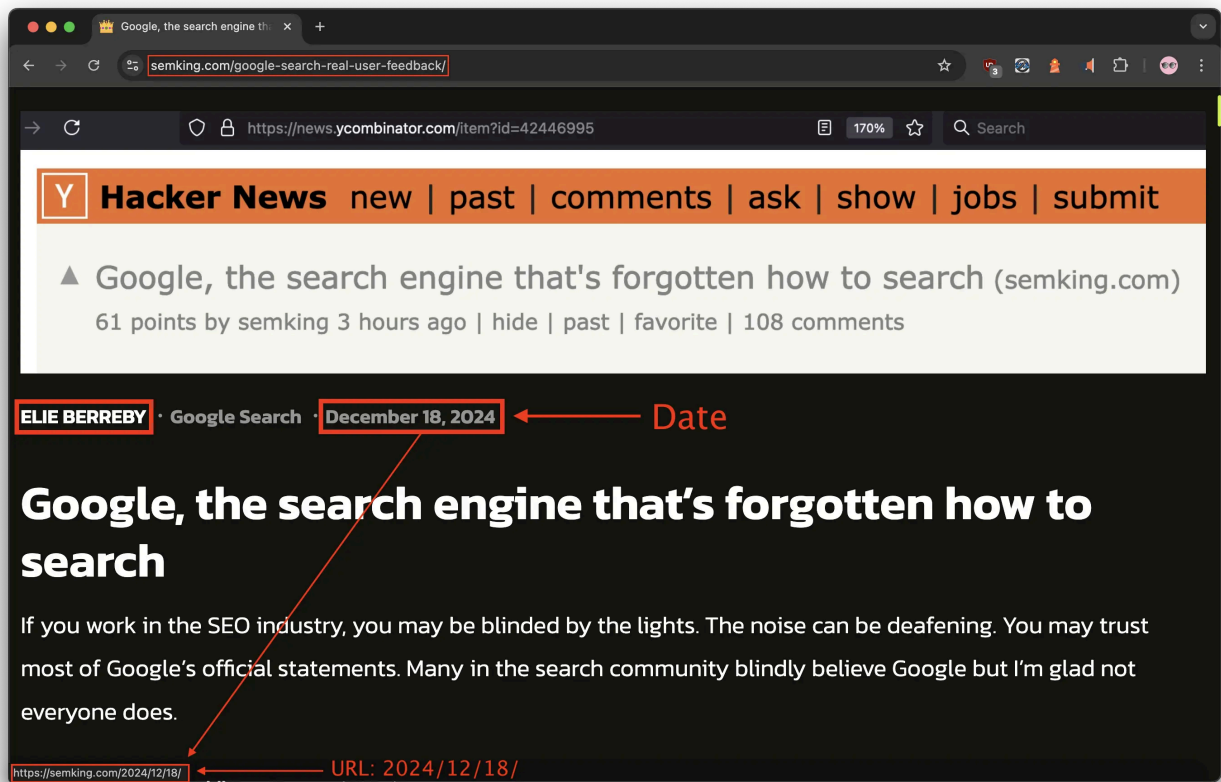
Broken Search Engines: Are We Losing the Web?

In 2025, we do not have functional search engines anymore. If we did, having advanced SEO knowledge would not be a prerequisite to being visible. Brands and people would focus on creating great content and they'd be rewarded by great rankings. Even and maybe especially when they'd ignore artificial optimizations.

But let's face it: the world we live in is fake. Welcome to our beautiful era in which artificial intelligence and natural stupidity combine to produce polished but tasteless content and messed-up search results. Independent thinkers and real creators are an endangered species. Okay, rant over. Back to my example.

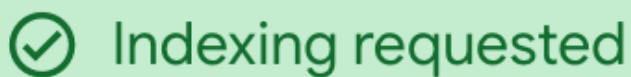
My useless indexing precautions

From a content perspective, everything was clear on my website.



Googlebot could “see” that my local publication date was December 18, 2024. The URL was available in my daily 2024/12/18 content (check the bottom URL in the screenshot). And yes, I updated my feature image and used a screenshot of my Hacker News post as a featured image because I loved your reactions! 😊

Everything should have been clear from the start... and you don't yet know that I had taken extra indexing precautions. I told you I was now opting for a passive do-nothing approach and that's true. But before I shared the URL of my article on Hacker News on December 18, 2024, I had made sure Google had indexed my content. How could I be sure?

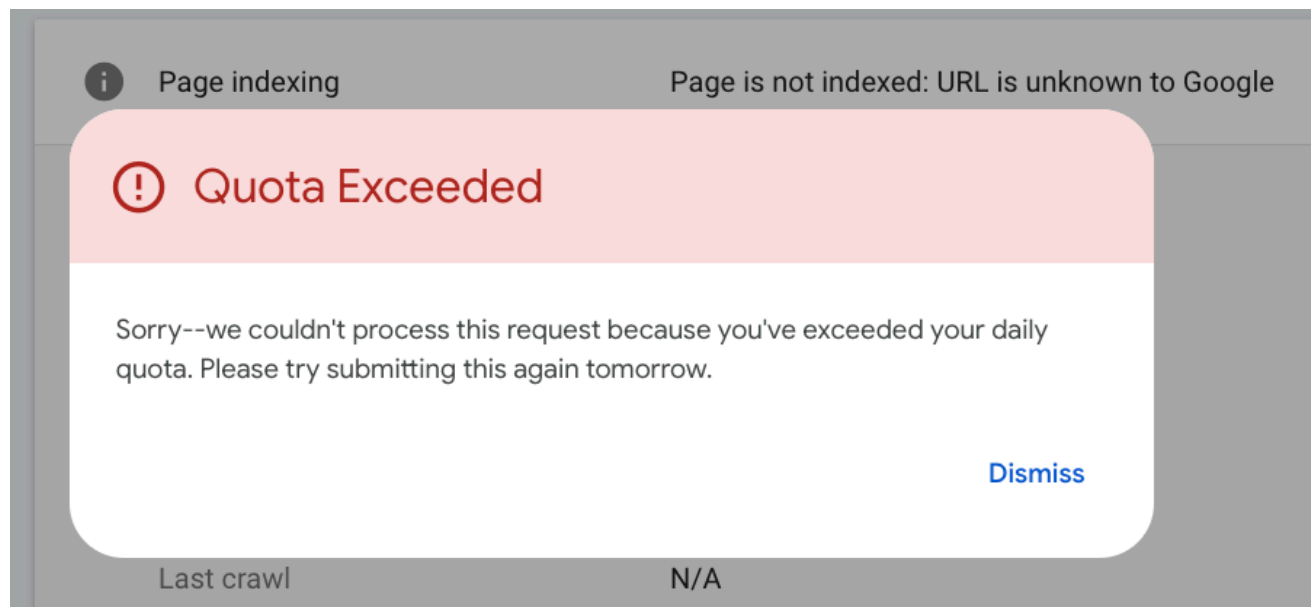


URL was added to a priority crawl queue. Submitting a page multiple times will not change its queue position or priority. [Learn more.](#)

[Dismiss](#)

I used a powerful method that allows you to speed up the indexing process for critical URLs. In my Google Search Console panel, I manually added the URL to a high-priority crawling queue and I manually ensured Google indeed indexed it by checking the index in real-time.

You can use the URL Inspector in your Google Search Console (GSC) panel to ensure your URL is in Google's index. But that's not enough because there's can be a latency between GSC and the live index. The data displayed in GSC often lags. Sure, when you test the live URL, it is supposed to be near real-time. But there's always a lag. Complicated topic, trust me.

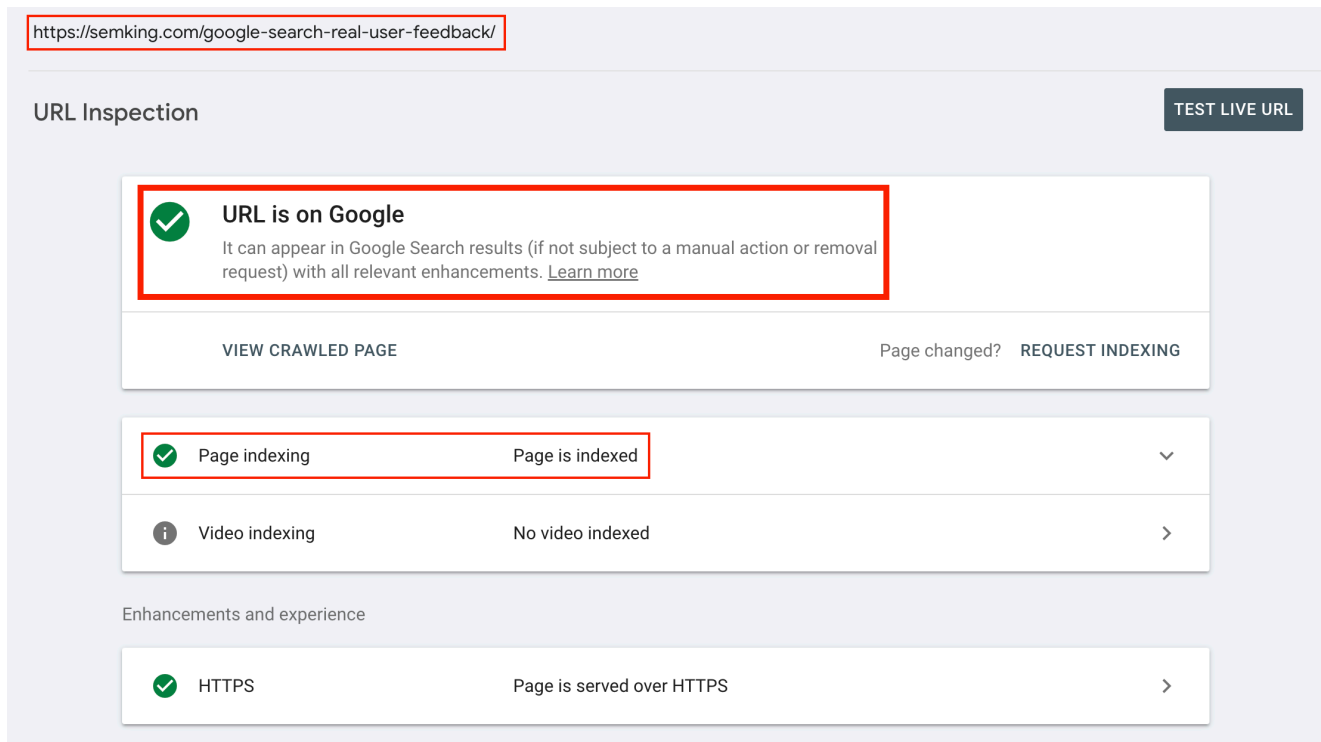


Pro tip: you should consider this method exclusively for critical URLs because the number of requests is limited to only a few per 24 hours.

If you overuse it, you'll get an exceeded quota error like this one (example taken from a client's website who requested urgent indexation of multiple URLs —and yes, I had to spread the work over several days).

Back to my HN example: I did things in the best possible way. I sent my exact URL to a high-priority crawl queue. I then manually verified the indexation of the exact URL of my article using the URL inspector in GSC. But knowing this was not enough, I also queried the live index. Some recommend using the "site:" operator but there are issues. Its output isn't always reliable. A simple trick? Just search for a unique sentence in your content. It could be one of your titles or a unique phrase in your text. If your URL appears, it is in the index!

Believe it or not, I did all of this to confirm indexation before sharing my URL on Hacker News.



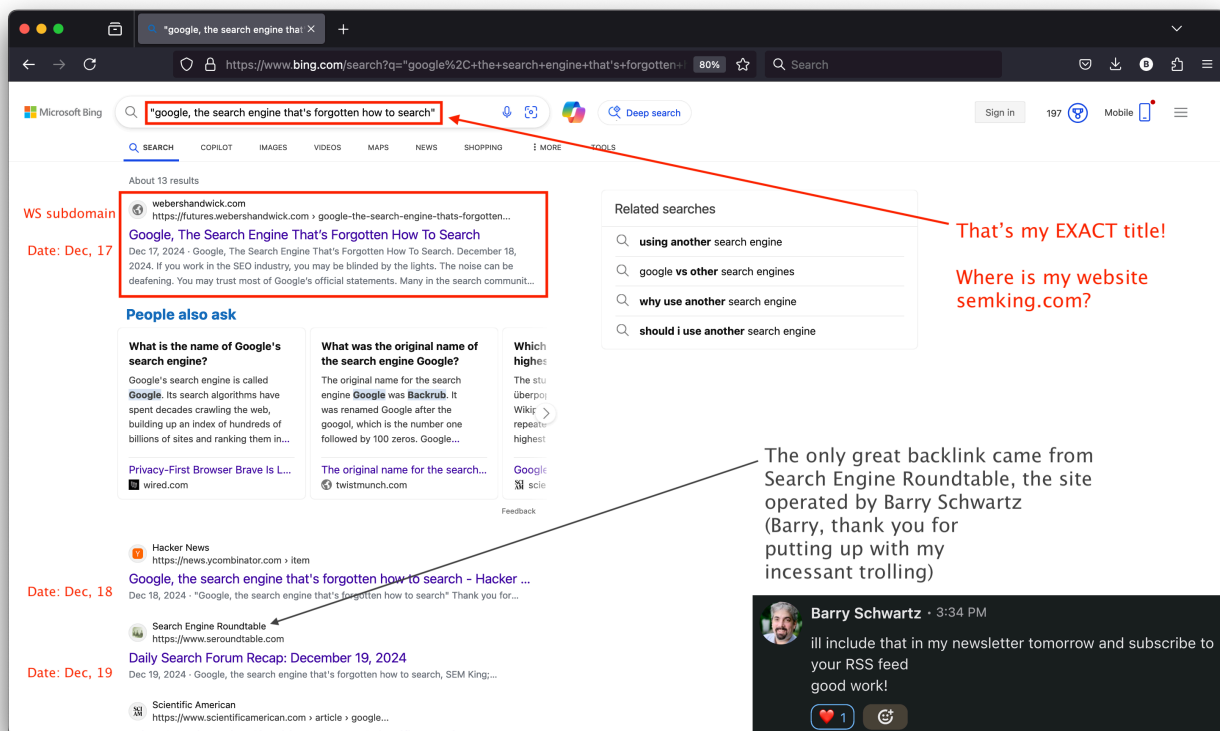
And yet, here we are today: my article does not even appear for its exact title in Google Search. Instead we have copycats hiding under the guise of syndication that rank for my article's title. But I can guarantee that those practices would stop if search engines were smarter —or just a little less stupid. None of the indexing precautions I took were enough.

A secret Google shadow-ban?

Some of you know that Google employees read my article. You may conclude that this is all happening because Google people found my content “problematic” and took some of action leading to a shadow-ban.

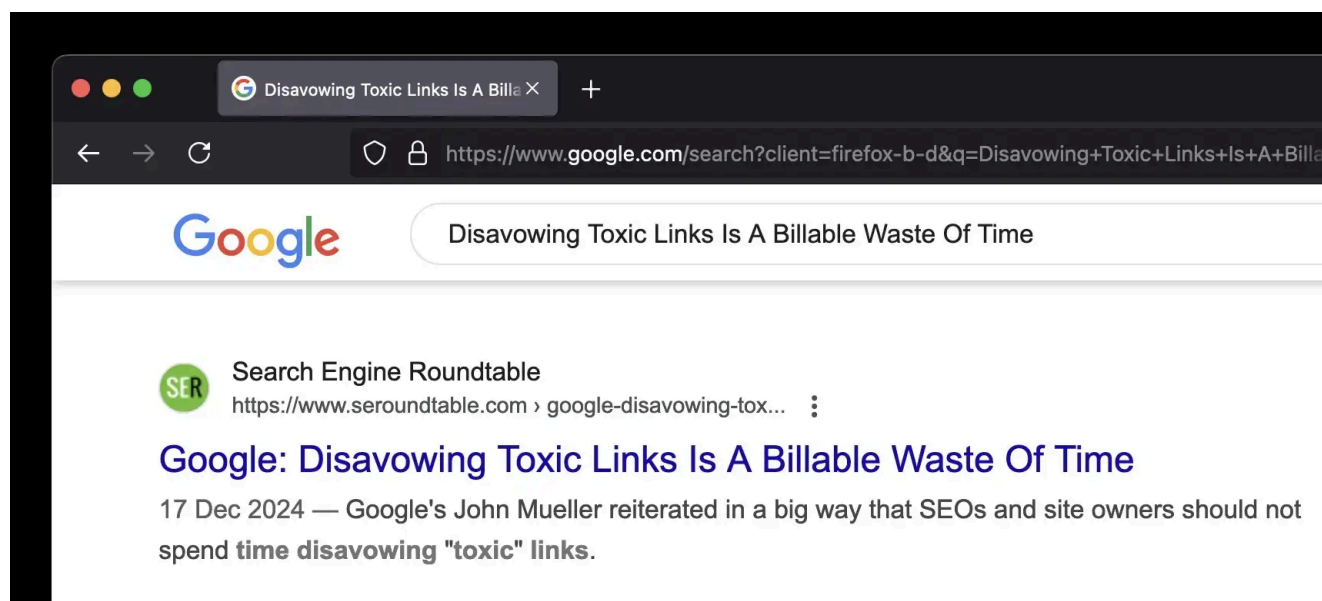
First, my reach is NOT important enough to deserve that level of attention. Secondly, hiding negative user feedback in the SERPs would be unacceptable —and the Streisand effect would hurt them. Thirdly, the Google/Alphabet employees who read my stuff secretly love it. We get along really well, especially with those who have blocked me everywhere! 🤔

Finally, accusing Google of unfair practices in this case would be stupid because I was also penalized on Bing. Here's a Bing screenshot I took hours before publishing this case study in January 2025.



This is an epic but global Search failure that also affects Bing. But let's again put ourselves in the virtual shoes of Googlebot. Many so-called "SEO experts" believe that any link is better than no link. That's why there's a whole "industry" to buy and sell links. The price of a great link can be mind-blowing! I've never once bought any link in 18+ years. Have I been kindly approached to discuss "placement" on the many I manage in exchange for money? Many times. But I never agreed.

That said, for website owners, buying links is a tactic that's worth trying. If it works, it works. If it does not work, it isn't a big deal. No penalties. After all, didn't Google itself just claim in December 2024 that "*Disavowing Toxic Links Was A Billable Waste Of Time*"? You can Google the Search Engine Roundtable article by Barry Schwartz—but you've got the idea.



The theory is therefore simple: bad links do not matter because Google is smart enough to simply ignore them and consider only the “good” links. But as you’ve seen, Google Search isn’t that smart. It is actually very stupid. Decent people have no idea how easy it is to manipulate SERPs for their benefit or to damage a competitor. Maybe I’ll write about negative SEO one day.

For now, I won’t argue with this Google statement but I’ll try to spark a debate that could trigger those who always believe Google’s official statements. There’s a difference between Googlebots (sorry, I couldn’t resist) and those who test and take every official statement with a giant bucket of salt. As far as I’m concerned, it would be unreasonable to blindly believe spokespersons paid by a company to publicly diffuse this company’s message to the masses.

For instance, I had a long conversation with a Google employee over many days regarding a highly technical topic. They were very knowledgeable and they pointed out potential mistakes I might have made regarding complex topics. And I actually believe they were more skilled than me, BUT...

The problem? Empirical data contradicted part of their speech. I’m sure they themselves believe the theory they were paid to diffuse but it turns out in the real world, things are different. And Google employees aren’t in the shoes of web publishers.

Google’s Ranking Signals: A Mystery Even to Its Own?

There’s one thing I can tell you with absolute certainty: no single Google employee knows every intricate detail about the hundreds of ranking signals, no matter how senior they are. There are ranking signals they do not understand. In listed companies, everything is highly compartmentalized. Nobody has access to everything, not even the C-level. You’d be surprised! Just imagine the damage if one disgruntled employee left while knowing everything.

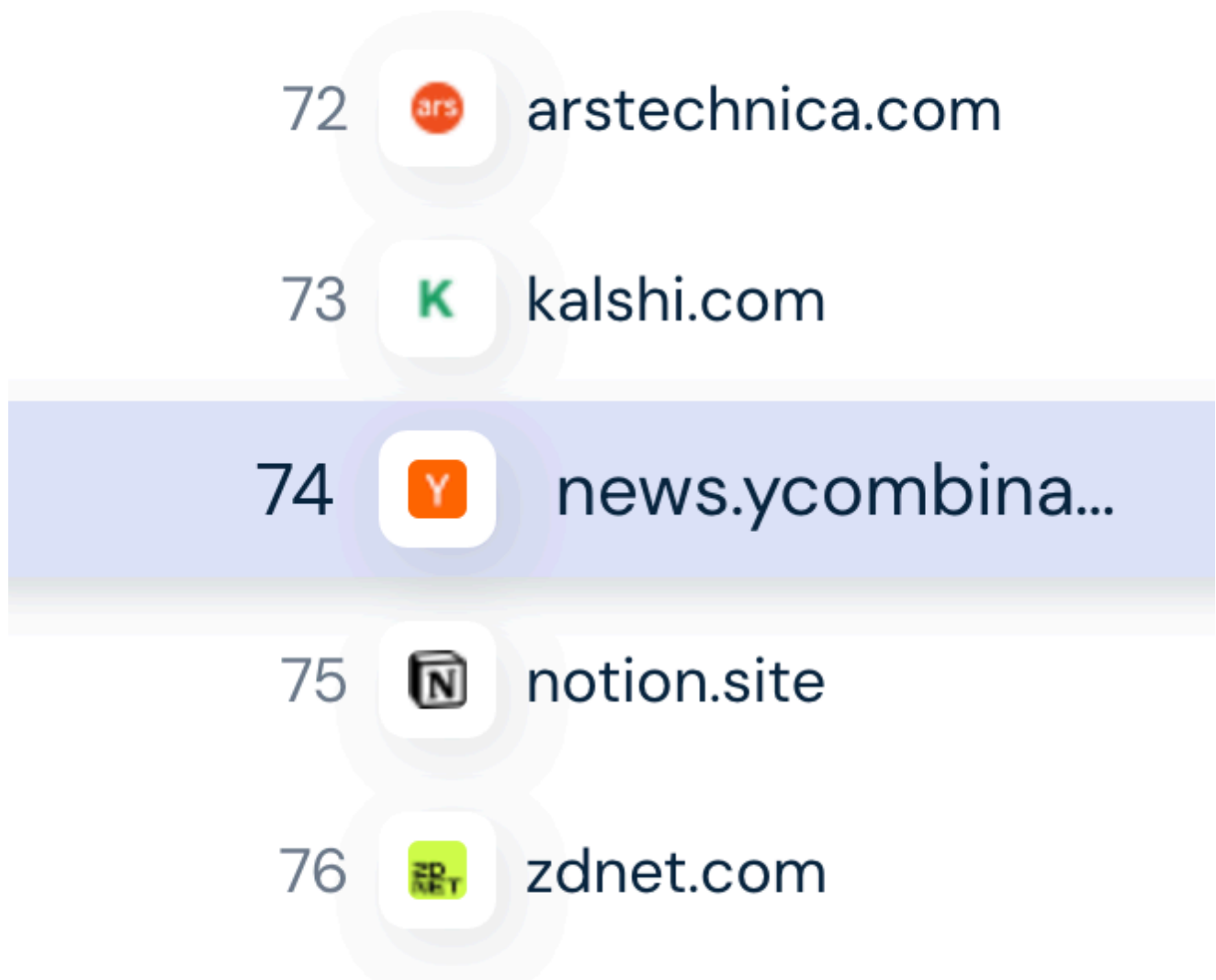
Only those who have spent considerable time and resource testing can be sure... but only for a short while because the search algorithms constantly change. Because change is the only constant in the Search industry, things remain interesting and challenging. The takeaway? Be skeptical of authority figures: some have no clue and some are paid to lie.

Backlinks: relationships you can’t choose

I see backlinks as relationships. And we can tell a lot about someone by observing their friends and who they surround themselves with. Problem: online, you don’t have much choice. Anyone can link to you without asking for your approval.

Do I believe low-quality inorganic links can damage new content creators and tiny brands such as startups? Yes because bad links can add a useless burden to websites that already suffer from a low quality score. Remember that small publishers have no established online reputation, they have everything to prove. Take my example. I only started posting on my

personal blog months ago so that's exactly me: a tiny content creator with everything to prove.



After sharing my URL on Hacker News, I noticed a surprising number of low-quality websites suddenly linking to my article. Many spammy backlinks came from an inorganic ecosystem that constantly scans Hacker News.

Because HN is immensely popular among a certain crowd, a few “tech bros” are abusing the system to make their web properties visible. After all, creating great content is incredibly time-consuming.

Their unspoken motto could be: why not just hijack whatever appears popular at any given time? Don't get me wrong, that's a global issue that any wildly popular website will face.

Hacker News: Where Clicks Turn into Hours

To explain why HN is a prime target of inorganic ecosystems, look at the Similarweb data for Hacker News / YCombinator during November 2024. More than 11 million visits and an average time on the HN website way above 4 minutes.



4+ minutes is great but that's not counting the time spent on the target URLs: that's just time spent on HN! Try to imagine the total visit duration and the user engagement that HN creates for target websites! Simply incredible!

Semrush, Ahrefs & online marketing tools

Online marketing analysis tools such as Semrush or Ahrefs extrapolate data using their own web spiders. Their web crawlers named SemrushBot and AhrefsBot try to mimic Googlebot but the data isn't always "accurate". The results can be problematic when users forget the extrapolated nature of the data. Nothing is 100% accurate but those tools can be helpful.

While we often have to deal with false positives and negatives, regarding velocity (the speed at which new signals are detected), those platforms often do a better job than Googlebot, especially for small sites.



At first glance, it may seem that Ahrefs' backlink checker got drunk, went home and concluded that nobody loved me. But that would not be accurate. Do you know what is going

on? Is the tool to blame?

- **semking.com** is my registered domain name.
- www.semking.com is a www subdomain, not the domain!

These are entirely different objects. In the past, creating a subdomain named “www” to specifically host website content was standard practice, but these days are gone. I don’t use a www subdomain because I don’t want to. Noise is useless.

Quick SEO tip: if your website isn’t set up to redirect users from one version to the other, search engines might see them as separate sites, potentially affecting crawling, indexing and ranking. And that’s exactly what is happening on my website: I added an .htaccess rule to redirect all traffic from the www subdomain to <https://semking.com/>

Technically, my website semking.com is hosted on a dedicated machine using a classic LAMP stack. For those unaware, **LAMP** is the acronym for the components in the stack:

- **Linux** (operating system),
- **Apache** (web server),
- **MySQL** (relational database) and
- **PHP** (server-side programming language).

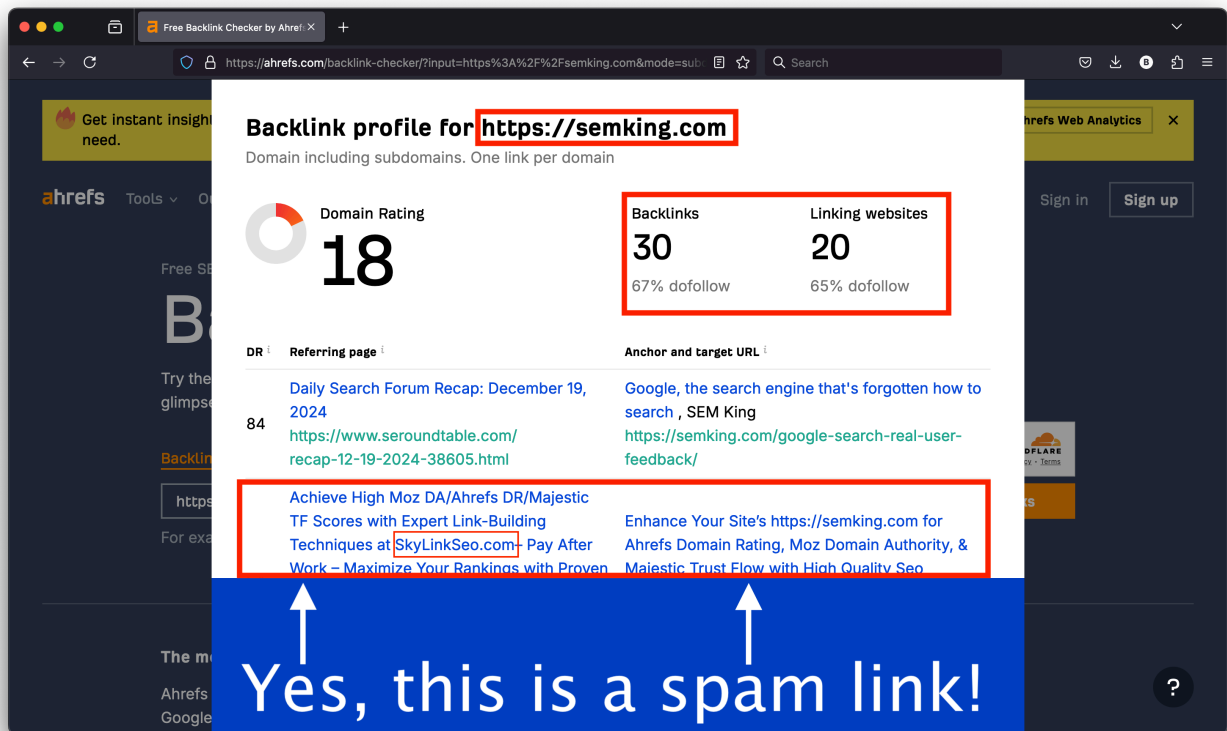
I love the Debian + Apache combo. Nginx is often superior but I decided to use Apache. For the databases, I’m actually using MariaDB, a fork made by the original developers of MySQL. It is faster and although I could do much better, I’m sensitive to web performance because of how it impacts the user experience.

Cloudflare, CDNs & crawl rate

Cloudflare is a fantastic Content Delivery Network (CDN) and I’m not using it just to optimize the user experience of my readers all over the world. Using a CDN can positively impact how Googlebot crawls your website. If Googlebot notices your server’s response slows down, it will throttle its crawl requests, meaning it will crawl at a slower rate. But when you’ve warmed up your CDN’s cache, Googlebot can crawl faster. And because the crawl budget is measured in seconds, the faster the spider can crawl during the pre-defined number of seconds, the more content will be crawled. And because JavaScript rendering eats your crawl budget, optimizing your rendering strategy makes sense. Cool topic but back to our case 😊

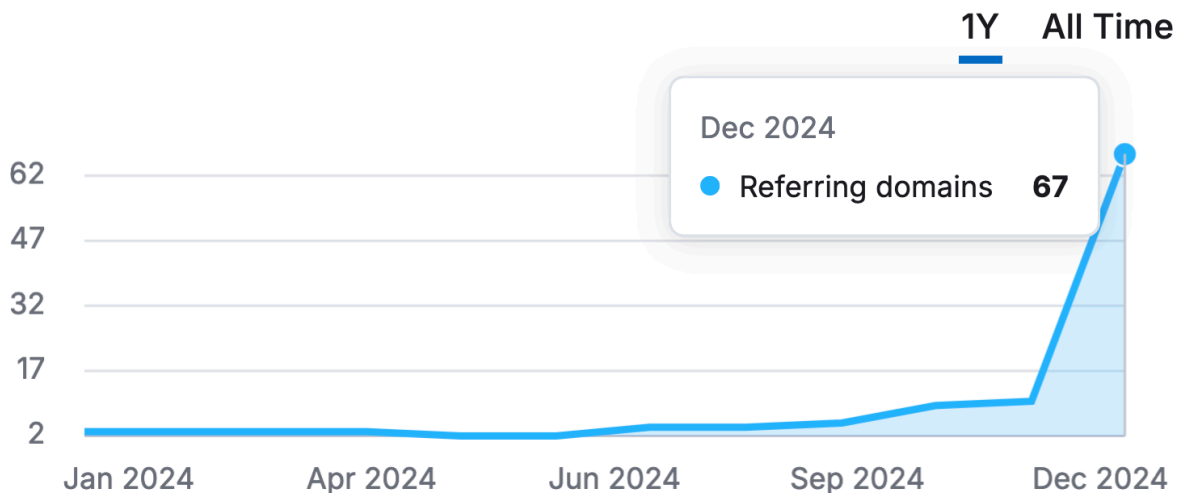
I created a redirect rule at the web server level (Apache) that redirects all traffic from the www.semking.com subdomain to my domain name (using the HTTPS protocol). Here’s the Ahrefs result when I enter my actual domain name, without the www subdomain. This tool detected 30 backlinks coming from 20 websites.

The first link is from Search Engine Roundtable but many are spammy...



Let's now look at the Semrush data showing the referring domain names that appeared immediately after my Hacker News publication. Those numbers keep going up as of early January 2025.

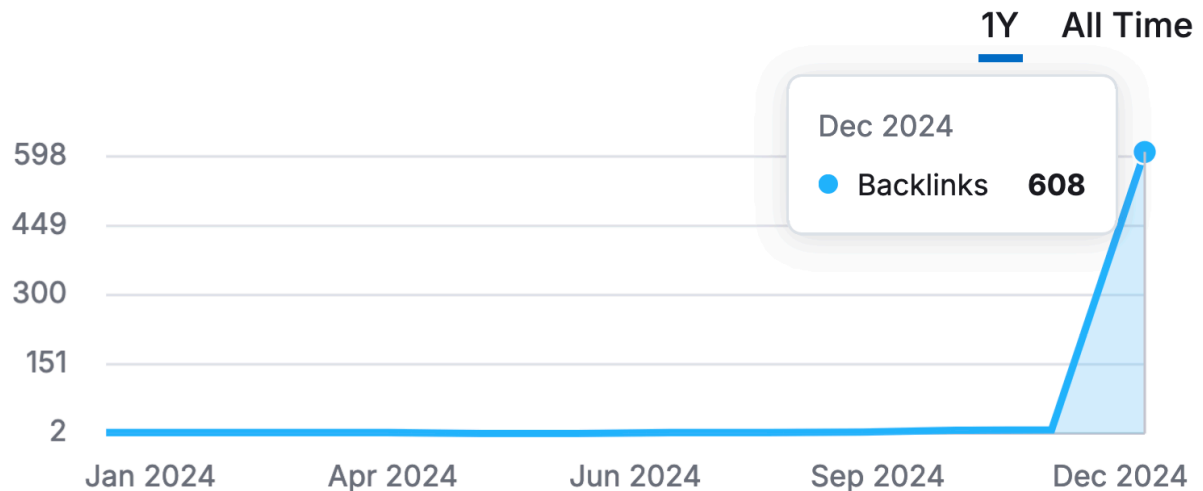
Referring Domains



Many “SEO experts” would love to see this jump in the number of referring domain names. I don't.

No digging required: when I look at the number of backlinks, it is evident that this didn't happen organically.

Backlinks



When I started writing this article, the ratio of backlinks to domains (the number of backlinks divided by the number of domains), was equal to approximately **9.07**. This data always evolves over time but an average of 9 backlinks per domain sends a negative SEO signal that search engines quickly notice.

For search engines, this abusive pattern is most likely the result of a manipulation attempt by the website “benefitting” from all those links (me). Apparently, you and I are the only ones who this isn’t true! 😊

Suppose you wanted to support my content. You would manually create a backlink from your website to mine. You’d most likely create ONE link to the article URL you liked and use a relevant anchor text. If you felt uninspired, you’d create an anchor text using my name: Elie Berreby. But you’d never create more than 9 backlinks to point to a single article.

Most of my new artificial backlinks were created programmatically and that pattern is systematically associated with low-quality. Did some websites mess up the average by creating an abusive number of backlinks?

The “top 5” referring sources created respectively:

- 390 backlinks from a single .org domain,
- 84 backlinks from a .com domain,
- 73 backlinks from a .com domain,
- 27 backlinks from a .com domain name and
- 16 backlinks from a .news domain name.

AS	Root Domain / Category	Backlinks	Country / IP
13	████████.org	390	78.46.████████
<div> <div> Authority Score is calculated based on: </div> <div> <div>Backlinks</div> <div>5.4K</div> </div> <div> <div>Referring Domains</div> <div>338</div> </div> <div> <div>Monthly Visits</div> <div>61</div> </div> <div> <div>Organic Traffic</div> <div>6</div> </div> <div> <div>Outbound Domains</div> <div>18.1K</div> </div> </div>			
		84	173.236.████████
		73	172.67.████████
		27	66.175.████████

When I checked the KPIs 11 days after the Hacker News publication, I could see 77 referring IP addresses with:

- 692 backlinks —an increase of +6,820%
- 70 referring domain names —an increase of +600%

Days later, we are now in 2025 and every single spam indicator has gone up! 😞

Two weeks after sharing on HN, here’s what Googlebot detected when looking at the “Links” section in Google Search Console. I’ve added the SemrushBot data because Google Search Console usually lags behind, especially when you monitor small websites like mine. Remember: small crawl budget! Ultra-low quality sources such as the most spammy .org website which has now created 390 backlinks from a single domain name may never show up in Google Search Console.

Google Search Console data displays the root domain name, even when websites link to you from subdomains. For instance, while Hacker News uses the **news[.]ycombinator[.]com** subdomain, GSC data will show you a link from **ycombinator[.]com**. You’ll have to dig a little to get more granular data.

The screenshot shows the Google Search Console interface for the domain **semking.com**. The left sidebar contains navigation links: Indexing, Experience, Enhancements, Security & Manual Actions, **Links** (highlighted), and Settings. The main content area is titled **Links > Top linking sites**. A blue box highlights the **SemrushBot** data summary:

SemrushBot	
Backlinks	692 +6,820%
Referring Domains	70 +600%
Referring IPs	77

To the right, a box shows **Total external links 37**, with **Googlebot** listed below. Below this is a table of **Top linking sites**:

Site
brutalist.report
ycombinator.com
reddit.com
[REDACTED].com
heyjk.com
telemetr.io
algolia.com

Two red annotations are present: "This backlink existed pre-HN publication." with an arrow pointing to the [REDACTED].com entry, and "All others shown are post-HN links." pointing to the other entries in the top linking sites table.

To follow the search engine guidelines and deserve great organic rankings, small web publishers must demonstrate expertise, experience, authoritativeness and trustworthiness. Google calls this E-E-A-T. The two sides of the E-E-A-T coin? Content and links! Usually, great content is rewarded by great links from reputable sources. So, what is a good link?

The anatomy of a good backlink

A great link comes from a trusted website, ideally in the same industry or at least from a web page with semantical proximity. The link has descriptive anchor text and is a dofollow, which means it kindly asks search engines to follow this link. Nofollow links are explicit requests: please do not follow this link. I'm linking but please don't share my online reputation with the target.

One subjective attribute of a great link? It was created organically. You didn't even ask for it. And of course, no automated system created many variations of this link. The value of multiple links from one source quickly decreases as the number of links increases. Past a certain number of links, it becomes not only useless but spammy and abusive.

Bad links: toxic relationships

But let me ask you: why would the opposite not be true? The universally accepted scenario is that lousy content gets no links. But is this really how things work? I'm asking a serious question: why would bad content not attract bad links? Why would the pair not deserve each other?

After all, the internet is by definition a network of interconnected entities: on the web, nobody exists independently. Any publisher knows that content alone isn't enough. Nobody sits in their virtual corner: on the internet, people and brands try to connect to other entities in many ways.

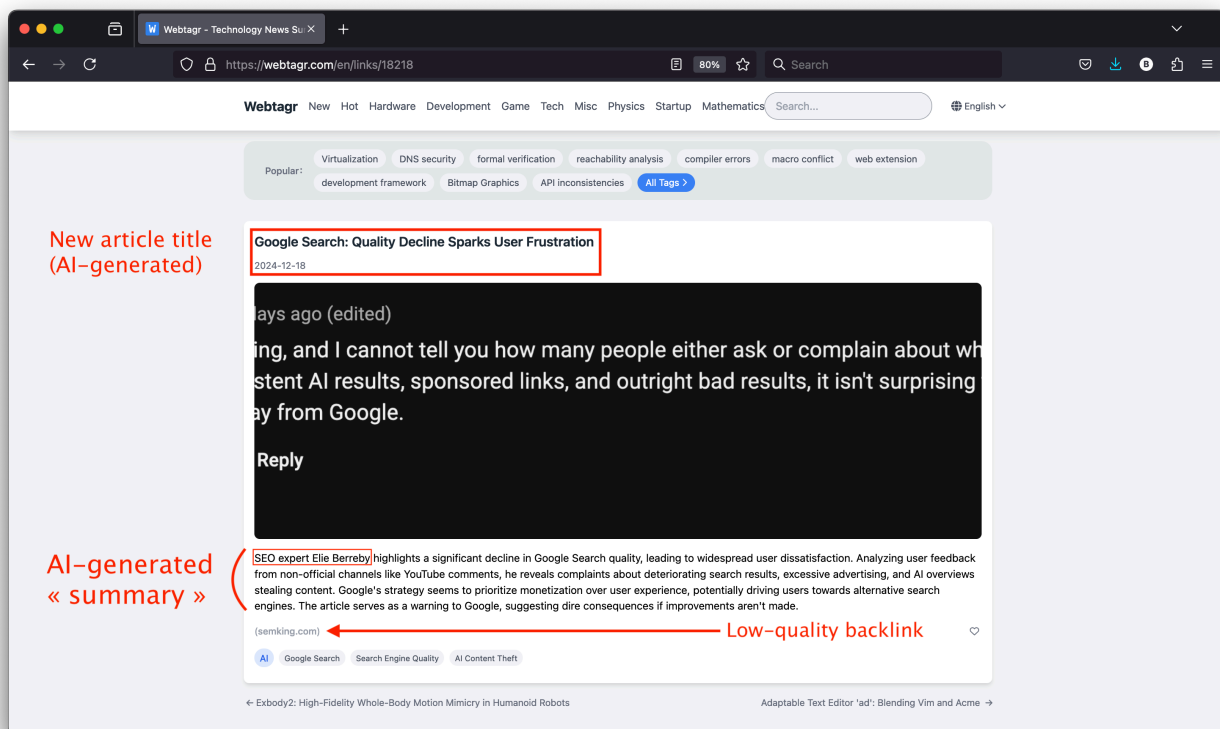
Remember that small web publishers are already dealing with two problems:

- Search engines give an unfair competitive advantage to large brands in SERPs
- AI answers scrape and regurgitate their content (content theft + engagement theft)

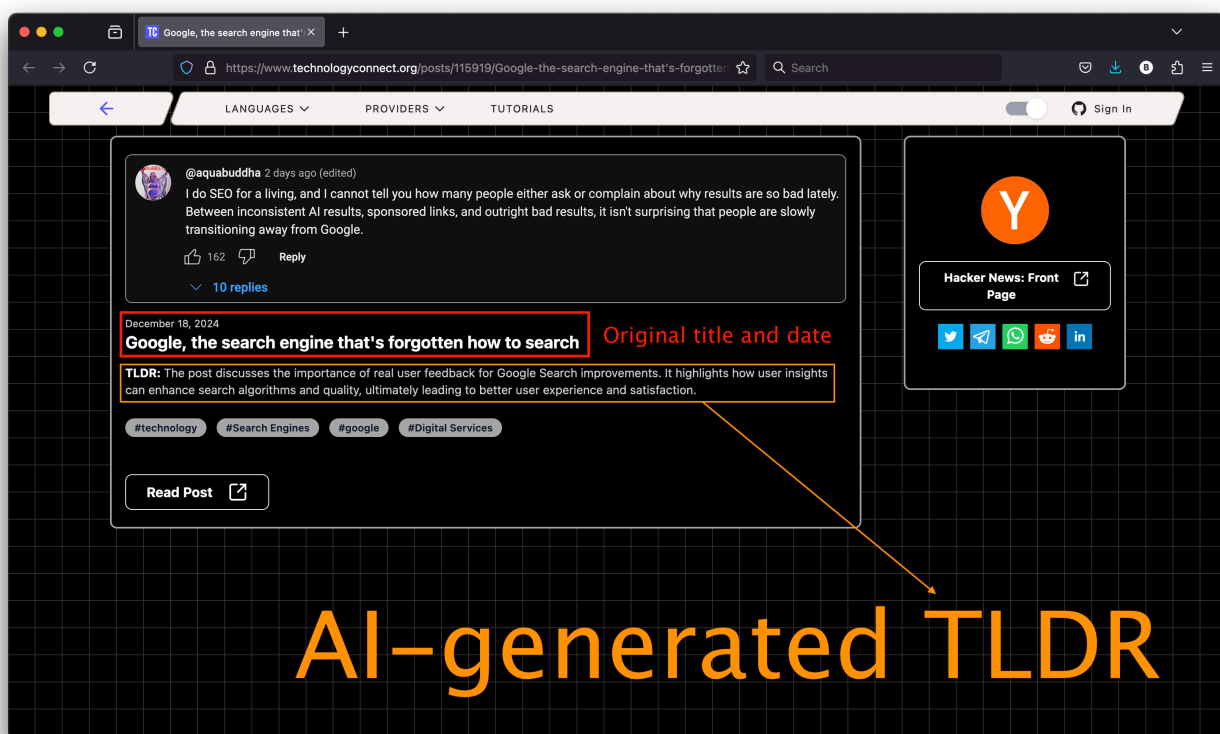
Drowning in Artificial Stupidity

Speaking of AI and content theft, the culprits aren't only search engines and their respective Large Language Models (LLMs). Our wannabe tech bro friends automatically fed my human-generated article to various LLMs. They then generated summaries and programmatically injected the short texts into their web pages to "create" new content. This process happens in seconds, without any human intervention.

Here's an example in which my article's title was transformed into "*Google Search: Quality Decline Sparks User Frustration*". This specific LLM decided to mention me as an "SEO expert", a title I would categorically refuse.



Here's another example. This time, they've kept my original title and generated a 2-line TLDR text.



Attribution: Where Does the Credit Lie?

My previous article initially published on MY website and later shared on Hacker News faces 2 paradoxical problems:

- **NO ATTRIBUTION:** my content was scraped and instantly republished on websites with a better Site Reputation and brand awareness in SERPs, all while failing to link back to my site or to credit me, the author.

- **SPAMMY ATTRIBUTION:** my article's URL has gotten hundreds of low-quality or even spammy links, a pattern that can end up destroying your credibility in search engines. While my website got backlinks from many low-quality sources after posting on HN, the "syndication" websites that scraped my text didn't get those awful links. After the HN event, guess who's backlink profile looks spammy and abusive? Mine, not theirs!

Search engines now think I stole my own content and quickly created hundreds of spammy links to my site. Lovely 😏

When Less Is More: Intentional Inaction

So, what am I going to do? As a senior growth advisor, I'd have many ways of "influencing" the SERPs. But I'll do absolutely nothing regarding my previous article—at least for a while! I'm going to stay completely passive and wait patiently like a wolf lurking in the shadows. I want to see if Google will correct its error or keep me shadow-banned. I'll pay close attention to what happens after sharing this case study with you.

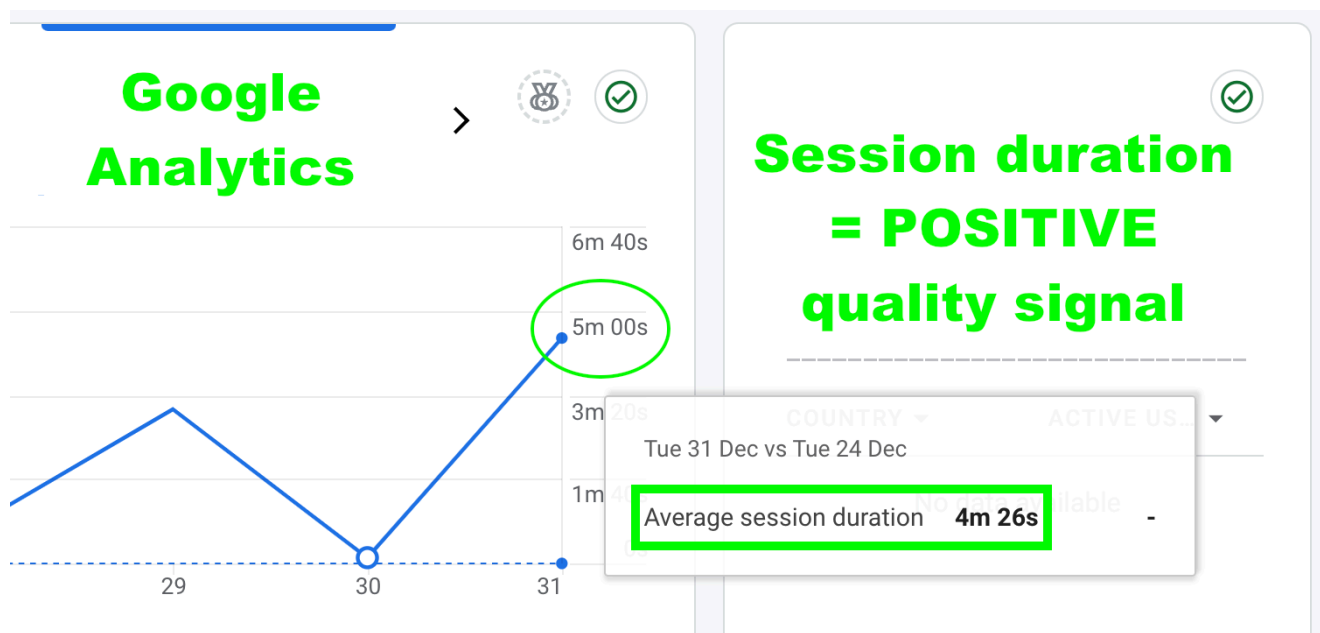
For the doubters, just search for the exact title of my article "*Google, the search engine that's forgotten how to search*" and you'll likely see unpopular LinkedIn and X posts ranking above my article. Actually my article URL is almost invisible. My site appears late in SERPs, after copycats and social media posts. And it only appears via its "Google Search Archives" (/category/google-search/ path): my article URL has been shadow-banned by search engines. The first readers should instantly Google this and take screenshots because I expect SERPs to change in the coming days and weeks.

People throw the "shadow-ban" expression carelessly but just test and Google it. And if you decide to share a screenshot with us, please consider adding your location because we all get different organic search engine results depending on our IPs. That's why your SERPs change when you use a VPN.

You can share your respective SERP screenshots in the Hacker News comments of this new article or in the comment sections of the LinkedIn or X posts I'll create prior to sharing this new URL on HN. I'll be grateful to anyone who does this, even if it is done to show the world how wrong I am. To me, what matters is reaching a credible conclusion, not being right.

Instrumentalizing Google Analytics?

Until now, I didn't use Google Analytics because I hate this tracking tool. But to continue experimenting, I've decided to configure GA4 before publishing this new case study and sharing this new URL on HN. Why? Again, there's an intentionality.



Until now, I was only using Google Search Console which exclusively processes organic traffic. Because Google Search Console only tracks organic traffic coming from Google Search, the signals associated with direct and referral traffic coming from Hacker News were not fully considered by Google. And Google could only “see” the time people spent on my website using the signals sent by the Chrome web browser. Today, this changes and while I might delete GA4 in the future, GA4 should send qualitative signals directly to Google because many people like you will spend time reading this new content.

Some readers spent 20+ minutes on my previous article: they watched my funny YouTube video, carefully looked at each screenshot, etc... But while the time they spent was a strong quality indicator, Google had not “seen” much previously. Using GA4 will create a precedent, a timeline that Google won’t be able to ignore.

How flawed is my analysis?

So, what’s the takeaway for you and me? Well, there are multiple other factors I didn’t mention that could have led to the situation I’ve exposed. Because of this, my entire analysis could be incorrect. Let’s be humble and keep in mind that no one knows for sure because many things are happening behind the scenes in an obscure way. Search engines are anything but transparent regarding how they rank content because they don’t want cheaters to abuse the system. Understandable.

That said, I did show you something incredible: how Google and Bing, the top 2 leading search engines worldwide, algorithmically penalized my article in a strikingly similar way. That’s very telling because Alphabet and Microsoft don’t use the same ranking algorithms in their search engines.

Google has developed a vast, intricate system of hundreds of ranking signals that have evolved over decades. Officially, Google’s ranking algorithms aim to deliver the most relevant, high-quality content to searchers like you and me based on a mix of relevance, authority, content quality and user behavior data. Epic failure, as we’ve seen.

Bing has its own unique set of algorithms and ranking methodologies. Bing's approach supposedly relies more heavily on machine learning, with claims that over 90% of its search results are influenced by AI following the deal with OpenAI.

How Google and Bing combat spam and low-quality content is supposed to be different, but we have seen very little difference in the output of their algorithms. Sure, my specific case is one example among many and while it should not be generalized, allow me to summarize: both search engines appear stupid and very easy to fool 😊

My Indexing Tips for Tiny Creators/Publishers

Here are some of the extra precautions I've taken to ensure I'll be seen as the original creator of this new article by Google Search and Bing Search. My overkill approach makes everything more time-consuming and complicated but I see this whole thing as an interesting search experience! 😊

1. Signing is Critical for AI search

Google Search has had problems identifying authorship for a while now. You may never have noticed this but it became clear to me after auditing multiple websites in 2024. And that's not all. Your strategy must consider "AI search". Here's why I recommend signing all your future articles with a signature such as

"© 2025 FirstName LastName + Brand + Date of publication + URL of the content."

Doing this systematically can slightly transform how conversational AI models answer over time. But your name must be repeatedly included in the body of your texts. Your name will no longer be independent from the text. Your name won't be found only in potentially hard to render sections of web pages. You probably never thought of this but did you know that, as of today, none of the popular "AI crawlers" render JavaScript apart from Googlebot?

What if your author's name and URL are hidden in JavaScript code that's not rendered?

Allow me to digress for a few paragraphs and in the end, everything will make sense. I want to explain a few things about tokenization because all language models break down text into smaller units called tokens. These tokens can be words or subwords. It all depends on the model's architecture and training data but that's the concept simplified.

Each small unit of broken-down text, each token, is converted into a vector (embedding) in a high-dimensional space. This embedding captures semantic and syntactic information about the token. During training, the language model learns to associate these embeddings with each other based on semantic similarity and co-occurrence.

Words that often appear together in similar contexts will have embeddings that are close in the vector space. That's co-occurrence. And words with similar meanings will tend to cluster together in this space. That's semantic similarity.

Some people use cosine similarity to analyze texts. It may sound complicated but everything about cosine similarity and angles is super simple. When two phrases are exactly the same, the cosine similarity is 1. When they have nothing in common, their cosine similarity is 0. When there's overlap/similarity, the cosine similarity is between 0 and 1.

But cosine similarity can actually range from -1 to 1, not just 0 to 1. A negative cosine similarity simply indicates that the phrases have opposite meanings in their vector space representation. With LLMs, we might not see negative values because the word embeddings are often designed to be in a positive quadrant of the vector space.

But let me ask you: what do you think is more important regarding your first name and last name? Semantic similarity or co-occurrence? Co-occurrence of course because most first and last names do not carry semantic meanings.

If your author's FirstName + LastName is common enough in the training data, your entire name might be tokenized as a single token. That happens for popular writers: their "FirstName LastName" appears frequently enough together in the training corpus to deserve a single token.

Unknown or unpopular authors require separate tokens. Their first and last names are tokenized separately. For example, I've been a nobody online for years because I love privacy. As a result, my name "Elie Berreby" would be tokenized into two tokens: ['Elie', 'Berreby'].

When names are used (books, citations, articles), the exact combination of first and last names (co-occurrence) helps in correctly attributing works to authors.

Let me ask you: what's worse than not existing as an author? Seeing popular LLMs confuse you with similarly named people, especially if they did bad things. I searched for several unpopular names and I got the craziest results. When the LLMs are "confused" they tend to hallucinate 😊

Authors with common names will benefit from disambiguation. When you search for information about or by an author, LLMs typically rely on exact or near-exact matches of names. Co-occurrence data helps because it ensures that the correct combination of names is matched. This reduces errors like confusing different authors with similar names.

In January 2025, none of the following crawlers can render JavaScript:

- 🚫 OpenAI (OAI-SearchBot, ChatGPT-User, GPTBot)
- 🚫 Anthropic (ClaudeBot)
- 🚫 Perplexity (PerplexityBot)

Of course, there are fail-safe measures to ensure you'll appear, but if a simple signature can optimize your existence in AI answers, why not do it? One line (your signature) in your text can help reduce future confusion.

Some people want to avoid ending up in AI datasets. Some do their best to get future mentions. Whatever your objectives are, remember that for now, JavaScript rendering is problematic for AI crawlers.



Kim Kuhlman, PhD · 1st

2d ...

B2B Content Marketing Strategist | SEO | AI Expert | Data-driven busine...

Then forgive the naive question, why do they bother with long robots.txt files?

Insightful · 💡 1 | Reply · 3 Replies

[See previous replies](#)



Elie Berreby **Author**

2d ...

Senior SEO n00b

Kim none of the popular "AI crawlers" render JavaScript:

- 🚫 OpenAI (OAI-SearchBot, ChatGPT-User, GPTBot)
- 🚫 Anthropic (ClaudeBot)
- 🚫 Meta (Meta-ExternalAgent)
- 🚫 ByteDance (Bytespider)
- 🚫 Perplexity (PerplexityBot)

The only one that fully renders JavaScript is Googlebot, which has implications regarding AI overviews.

But it seems we have different topics in mind?

Bad bots are aggressive web crawlers that ignore the directives in the robots.txt file. You are right, robots.txt directives are suggestions, not rules. We assume most respectable bots would abide by those directives.

That said, I often bypass and ignore those directives myself when conducting technical SEO audits because this can help identify issues. Bad bots do it for different reasons (usually to access confidential content).

Insightful · 🧠👍 2 | Reply

Signing your content will make your name part of the core text. You'll end up in the datasets of LLMs such as the ones used by Google's AI overviews, ChatGPT, Gemini, Claude. Those crawlers aren't designed to ignore core/body text.

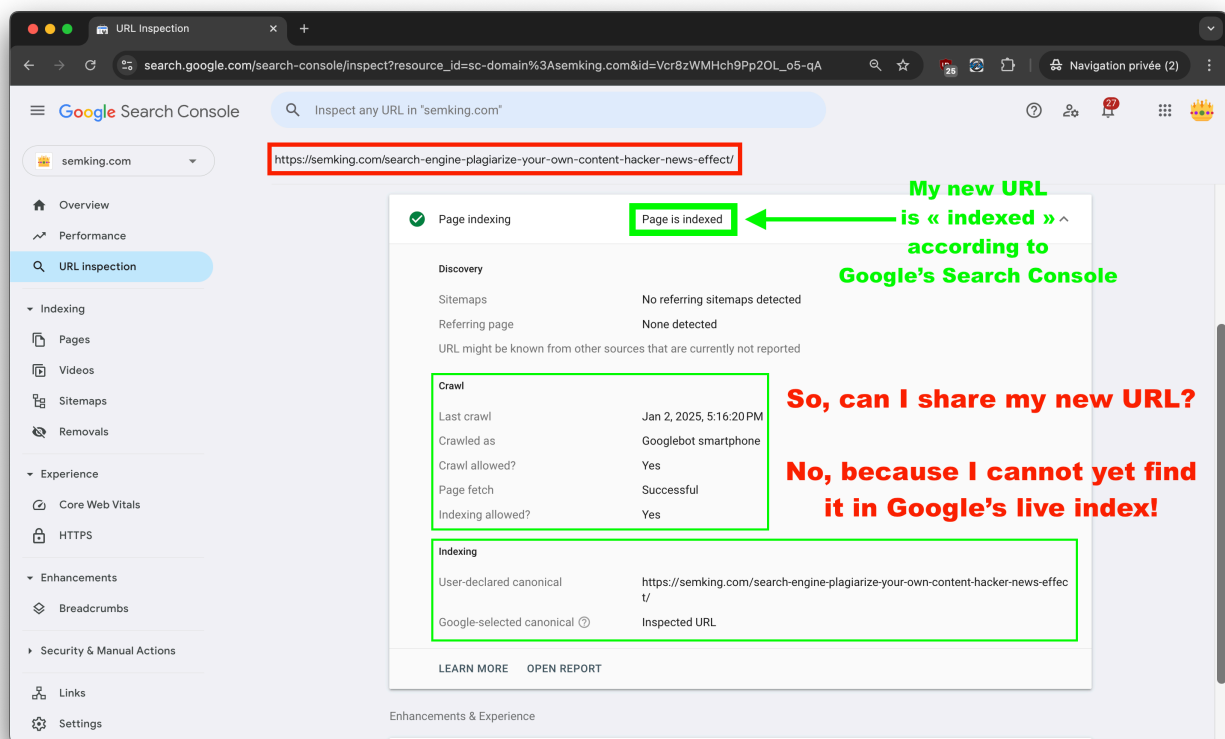
Nothing is organic anymore anyway so you might as well be smart and take advantage of this as an author/creator. I'm giving you a potential edge and I plan to use this myself from now on.

2. Indexing timeline

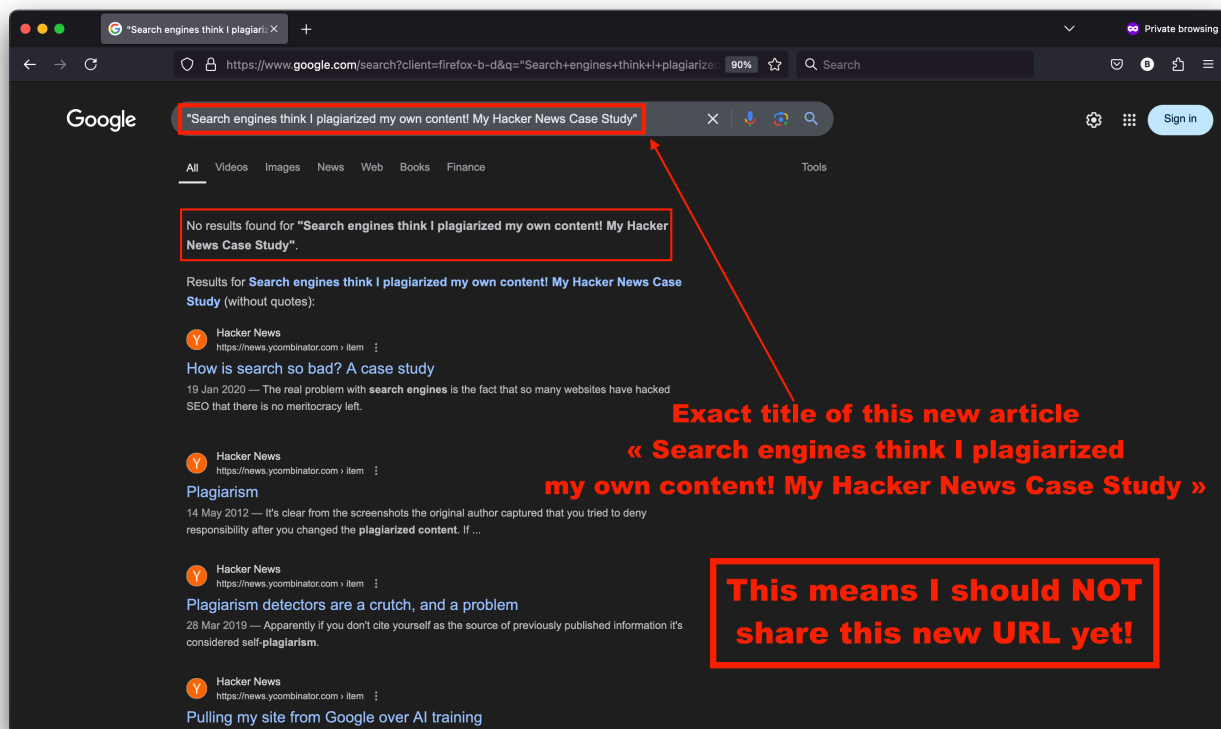
If you are a small creator who publishes on an obscure website that does not yet have a strong brand reputation, do NOT immediately share newly created content on popular websites such as Hacker News. Remember: the crawl budget allocated to your website and your site quality score are likely low. If Googlebot does not crawl AND index your content, do NOT share it because the inorganic ecosystem I highlighted could damage you. Timing matters.

As a web publisher, you may think waiting 1-2 hours is enough. It may or may not be enough. You'll have to manually TEST to ensure your content is indexed by Google and Bing under the correct URL and not a canonicalized version. If you cannot find your new URL and some exact sentences when searching, do NOT share it anywhere!

Does this look like a false positive regarding indexing coming directly from Google Search Console?



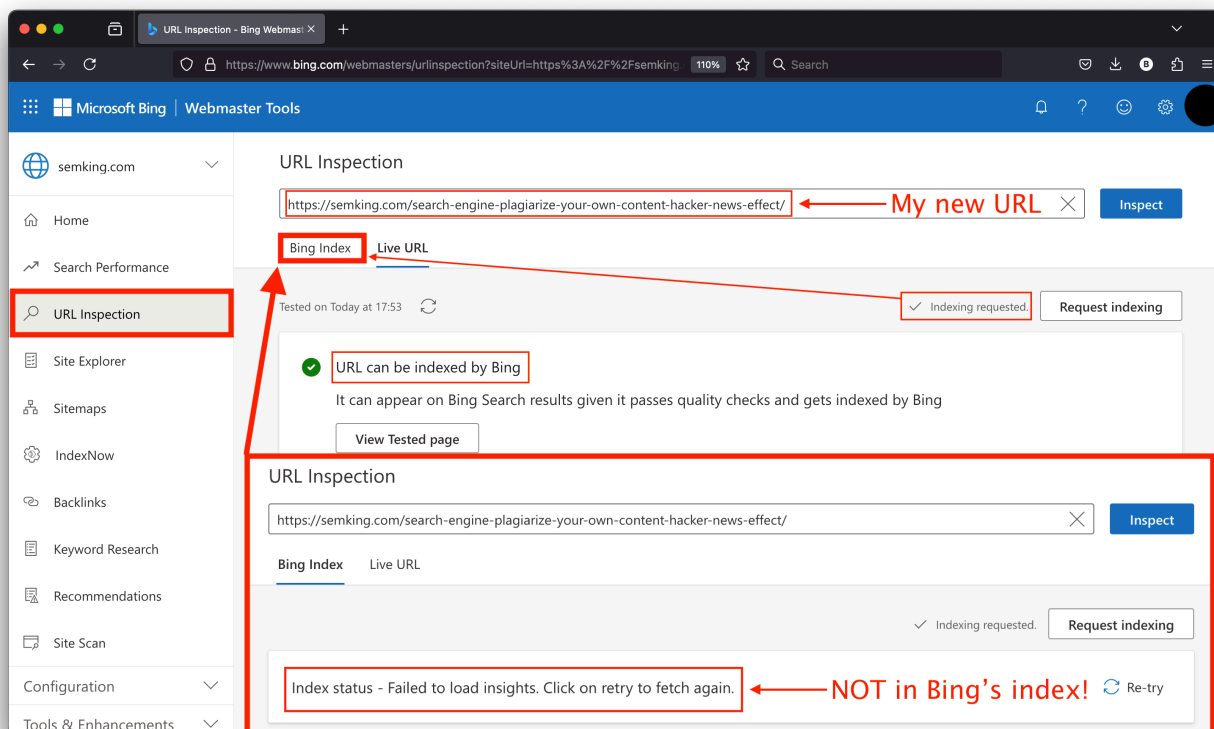
Google claimed the URL of this new article was indexed. But because I could not find it in Google's live index, I decided to wait. I wanted to see the page in the index before sharing it. You should combine the URL inspector in Google Search Console, the "site:" operator in Google Search and search for exact sentences in quotes in Google and Bing before sharing your URL. I queried the live index repeatedly for about 3+ hours. No result!



We want to see your page in the live index. That's the only definite indexing proof!

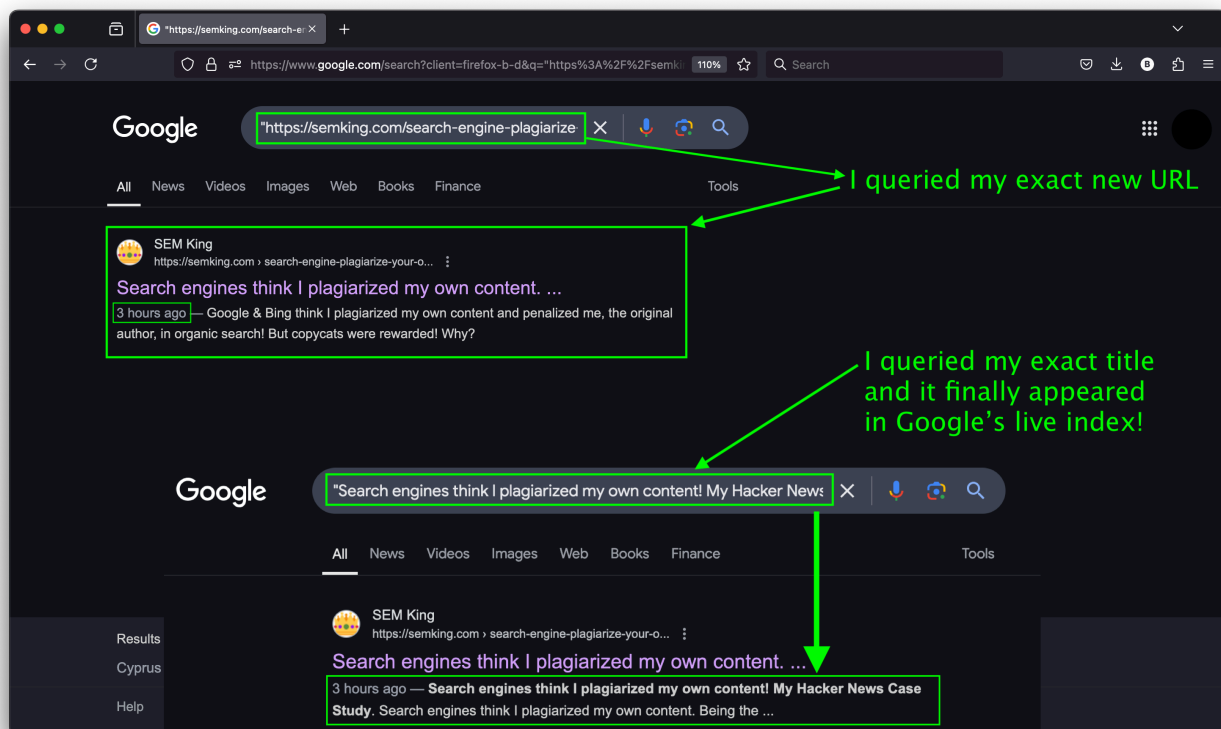
While you are waiting, why not manually add your new URL to Microsoft Bing's Webmaster Tools?

The process is very similar to the one I've highlighted in Google Search Console. Here again, you'll have to request indexing MANUALLY. You'll be limited to 100 URLs per day, a lot more than the quota in GSC. You'll have to check the Bing index in the Webmaster Tools AND in Bing's live index. Same logic if being indexed in Bing is important for you: if you cannot find your URL or an exact unique sentence in your article, wait before sharing your new URL!

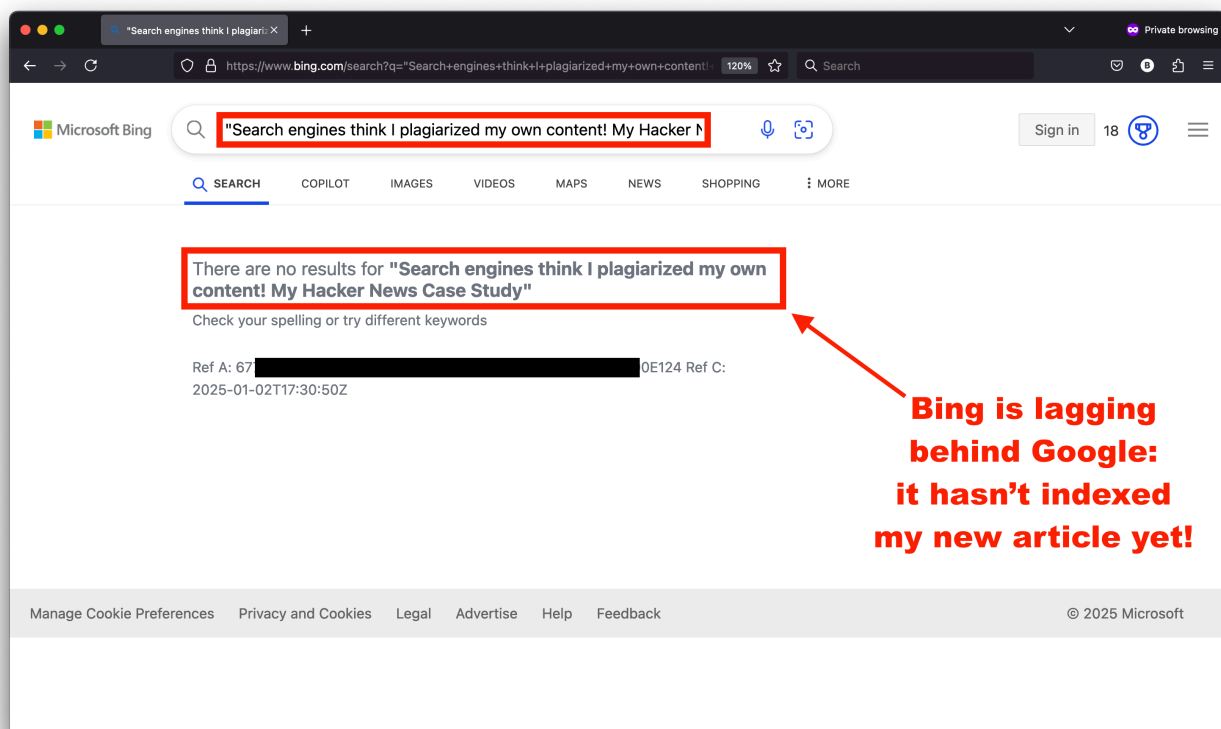


What about large websites? Well, it is always worth taking indexing precautions, but Googlebot will do most of the work for you if you are a trusted web publisher with a high crawl budget!

I recommend this indexing strategy mainly for tiny web publishers and creators because I want to protect them from content and engagement theft. I decided to query the live Google index one last time and suddenly, my URL and exact title appeared with a publication date of “3 hours ago”. I can now share this new article with the world! Hacker News, here I come! 😊



Bing hasn't indexed this new article yet but I'm done waiting. I will share the URL with everyone anyway and you'll let me know if you'd like me to create a follow-up article of this follow-up! 😊



3. Watermarking

Watermark your images. Content scraping is endemic and while you cannot really watermark your text, you can tag your images. That's what I did for some screenshots in this

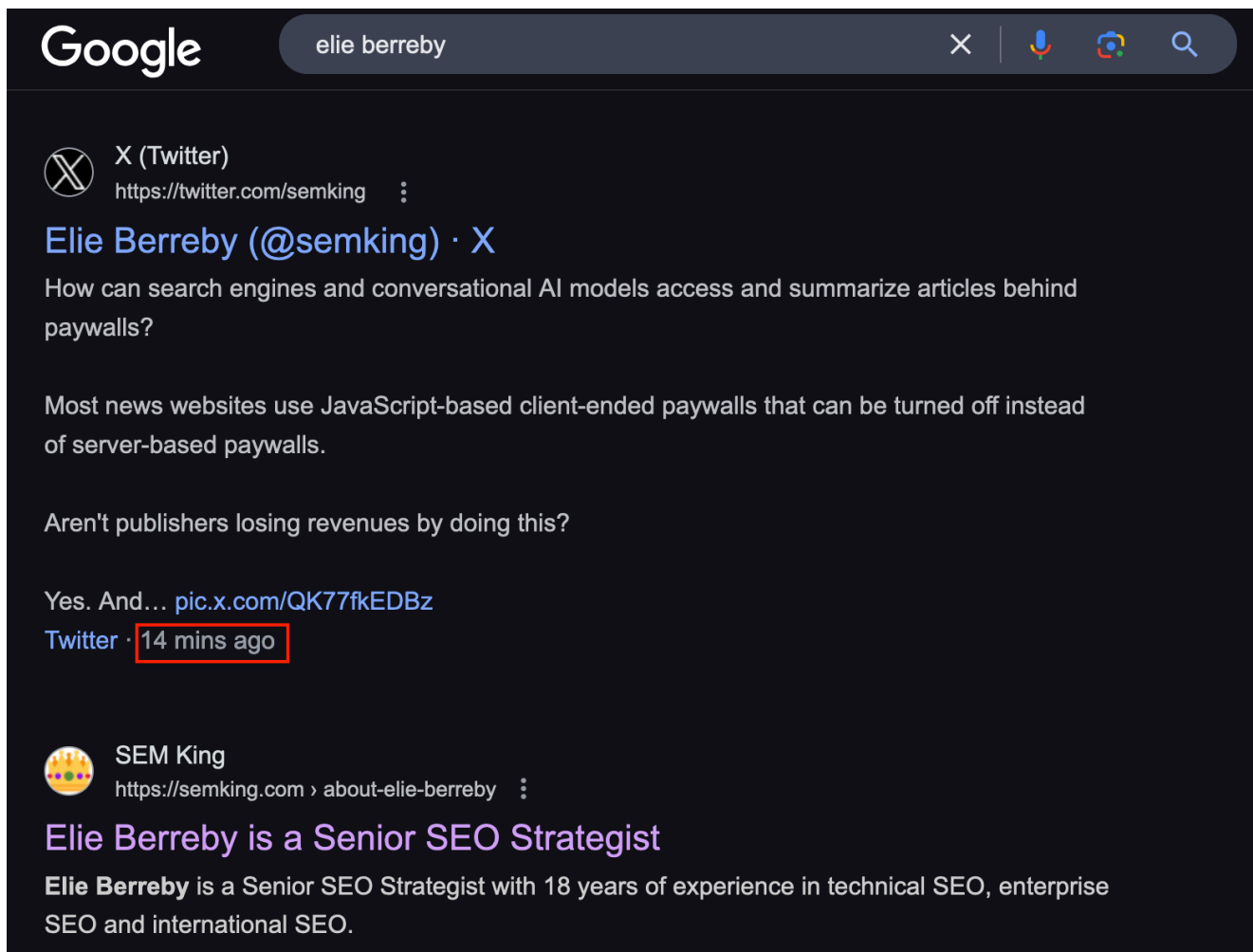
article. More time spent but content theft seems to be a new sport at this stage so better be safe than sorry. Most new generation LLMs are multi-modal, meaning they process images and videos.

You have different watermarking options depending on your site's size and the JavaScript framework or Content Management System used in production. The popular WordPress CMS has multiple open source watermarking plugins. If you only have a handful of images, you could also do it manually, at an operating system level (nerds know) or using design platforms such as Canva. Many options, do your own research and remember: if you do not protect your creations, they will be stolen.

4. Talking to the press or Sharing on your own?

If you are a brand, I suggest talking to the press and asking for links. Corny, but try to get backlinks from trusted sources because once you have many strong backlinks, spammy links won't negatively impact you. Remember: search engines look at your virtual "friends". If you only have weak links, you are the weak link!

I understand digital PR isn't accessible to most: it is time-consuming and costly. So here's another trick that could help before sharing great content on sites such as Hacker News. I've noticed Googlebot recently started indexing X posts very quickly.



The screenshot shows a Google search interface with the query 'elie berreby'. The results are displayed on a dark background. The first result is a tweet from Elie Berreby (@semking) asking how search engines and AI models can access content behind paywalls. The tweet includes a link to a picture on X and was posted 14 minutes ago. The second result is a bio for Elie Berreby, identifying him as a Senior SEO Strategist with 18 years of experience.

Google elie berreby

X (Twitter)
<https://twitter.com/semking>

Elie Berreby (@semking) · X

How can search engines and conversational AI models access and summarize articles behind paywalls?

Most news websites use JavaScript-based client-ended paywalls that can be turned off instead of server-based paywalls.

Aren't publishers losing revenues by doing this?

Yes. And... pic.x.com/QK77fkEDBz

Twitter · 14 mins ago

SEM King
<https://semking.com/about-elie-berreby>

Elie Berreby is a Senior SEO Strategist

Elie Berreby is a Senior SEO Strategist with 18 years of experience in technical SEO, enterprise SEO and international SEO.

How long will this last? I don't know. My guess is that there's some kind of deal between X and Alphabet but it could be a temporary rogue attempt from Google to crawl/scrape many X posts to allow Gemini (Google's multi-modal LLM) to compete with the growing user-generated content in xAI's dataset (yes: to compete with Grok).

My X profile is probably one of the worst you've ever seen and yet, whenever I've published anything on X in the past days, my short posts got indexed almost instantly —despite adding very little value.

Popularity does not really matter here: some social posts are being indexed simply because they exist...

Yes, I did try to make the search engine's job easier by having a basic Structured Data strategy that helps connect the dots between entities. The Schema Markup Validator shows you the result of my experimental JSON-LD code.

The screenshot shows the Schema Markup Validator interface. On the left, the JSON-LD code is displayed with line numbers 158 to 190. A green box highlights the code, and a green arrow points from the text "JSON-LD code" to it. The code defines a Person schema with properties like name, jobTitle, url, image, description, sameAs, and knowsLanguage. On the right, the structured data output is shown as a table. A green box highlights the table, and a green arrow points from the text "JSON-LD code" to it. The table has columns for the property name and its value. The values are: @type: Person, name: Elie Berreby, jobTitle: Senior SEO Strategist, url: https://semking.com/, image: https://semking.com/Elie-Berreby.webp, description: Elie Berreby is a Senior SEO Strategist with 18 years of experience in developing and implementing successful programmatic SEO and international SEO strategies. Elie Berreby has a proven track record in organic traffic optimization, website visibility, and Conversion Rate Optimization (CRO)., sameAs: https://www.linkedin.com/in/organic-growth/, https://x.com/semking, https://www.youtube.com/@technical-seo, https://www.techrepublic.com/article/google-changes-search-eu-dma/, https://www.business.com/articles/5-digital-marketing-certifications-you-should-know-in-2024/, knowsLanguage: English, French.

Property	Value
@type	Person
name	Elie Berreby
jobTitle	Senior SEO Strategist
url	https://semking.com/
image	https://semking.com/Elie-Berreby.webp
description	Elie Berreby is a Senior SEO Strategist with 18 years of experience in developing and implementing successful programmatic SEO and international SEO strategies. Elie Berreby has a proven track record in organic traffic optimization, website visibility, and Conversion Rate Optimization (CRO).
sameAs	https://www.linkedin.com/in/organic-growth/
sameAs	https://x.com/semking
sameAs	https://www.youtube.com/@technical-seo
sameAs	https://www.techrepublic.com/article/google-changes-search-eu-dma/
sameAs	https://www.business.com/articles/5-digital-marketing-certifications-you-should-know-in-2024/
knowsLanguage	English, French

Many would suggest promoting your content on Reddit because we know there's a business deal between Google and Reddit, Inc. We've all seen how often Reddit URLs appear in Google's organic search results. You'll decide if that's a good move for your brand. I won't suggest this approach because I've noticed too much censorship and SERP manipulation using this platform. But it's up to you.

Another idea after publishing your article on your website: quickly create a LinkedIn article (make sure you've allowed search engines to index your LI articles) and create a link to the article on your website from your LinkedIn article. With 1+ billion users and many link pointing to linkedin[.]com, LinkedIn's Site Reputation is light-years ahead of most websites.

In other words, ensuring your original URL on YOUR website was indexed in search engines before creating a LI article is essential!

5. Sending quality signals to Google

As much as I hate Google Analytics, if you know you are about to publish a viral or high-quality piece of content on your website, I suggest you configure GA4 a few days prior to publishing. That way, Google will start collecting quality signals and you'll create a demonstratable chronology/timeline.

Trust me, it would not make sense for search engines to ignore thousands of people spending 10+ minutes reading your great content! But as I've explained, if the traffic source isn't Google Search directly, Google Search Console will never track this.

Google Search uses predictive analytics to... predict the users' Click Trough Rate, the session time and many other metrics. If anything differs significantly from the expected pattern, positively or negatively, their system will re-consider everything and give you more or less organic visibility.

The longer your visitors stay on your site, the more your content appears enjoyable and satisfying to users! Collect those phenomenal quality signals because Google cannot afford to ignore them. And although the Chrome web browser directly sends invisible signals to Google, that's not enough. You want to help Google collect positive signals regarding user interaction. That's why GA4 can help. I'd suggest connecting your Google Analytics and Google Search Console panel.

6. Thinking of others: accessibility

Much like Google, your audience may suffer from visual problems. After I shared my previous article on HN, a reader named Jon wrote the following email to me.

On your website you mention "If you work in the SEO industry, you may be blinded by the lights"

Well as it turns out I am blinded by the darkness instead. I'm 54 years old and my eyes do not work the same as yours or as mine did a few years ago. Yes, that is a surprise to me too although back in the day everything IT was dark – white/grey/green/orange on a black terminal background.

I'd be really grateful if your website would either notice that my browser says please be darkly coloured typeface on a light background and do so or provide an obvious toggle.

Accessibility is hard but I hope you don't have to become somewhat decrepit yourself to take action! Oh, sorry – forgot to mention: great content – thank you.

Jon

Jon has spent many years in front of terminals and he now has visual problems. I suffer from the opposite problem: working on a light mode is very painful. Jon asked for a light mode. I cannot work on this right now but following his email, I tweaked the CSS of my WordPress theme to increase contrast and improve readability.

For now, I'll keep the dark mode but I hope that will help... and Jon's email gave me an idea. Remember point 4 in which I suggested sharing your content on social platforms such as LinkedIn and X? Well, those websites both offer light and dark modes. Creating articles on X or LI can not only help with indexing, this can solve the accessibility problem if your website does not offer an option to switch between modes.

And there's another idea I wanted to share following Jon's email. By adding a downloadable PDF with your article, you are not only helping readers who might prefer a white background: you may also help crawlers. You see, search engines love PDFs. I might explain why one day but in the meantime, just add PDFs to your important articles. And don't forget to name your PDF using your first and last names 😊

[Downloadable PDF \(28.5 Mb\): Search engines think I plagiarized my own content! My Hacker News Case Study \(Dark Text on White Background\)](#)

7. Scaling & Expanding

If you have a multilingual website using a gTLD domain extension (or a portfolio of ccTLDs) with authors publishing in multiple languages, my previous recommendations apply at a local level. Check my short [backlink profile dilution](#) post.

I hope this content was helpful. Writing this took a lot longer than expected. If you want to support me, you can upvote this on Hacker News, create organic backlinks from your website to mine and share this content on social media. You can find me on LinkedIn and on X @semking.



I'm going to look for a new full-time freelance role in 2025. My company is based in Europe but I want to work in Eastern Time because a natural inclination makes me more alert, energetic, and productive in the evening. I'd love to understand my chronotype but I don't. Please reach out if you can help me find a challenging SEO or organic growth role!

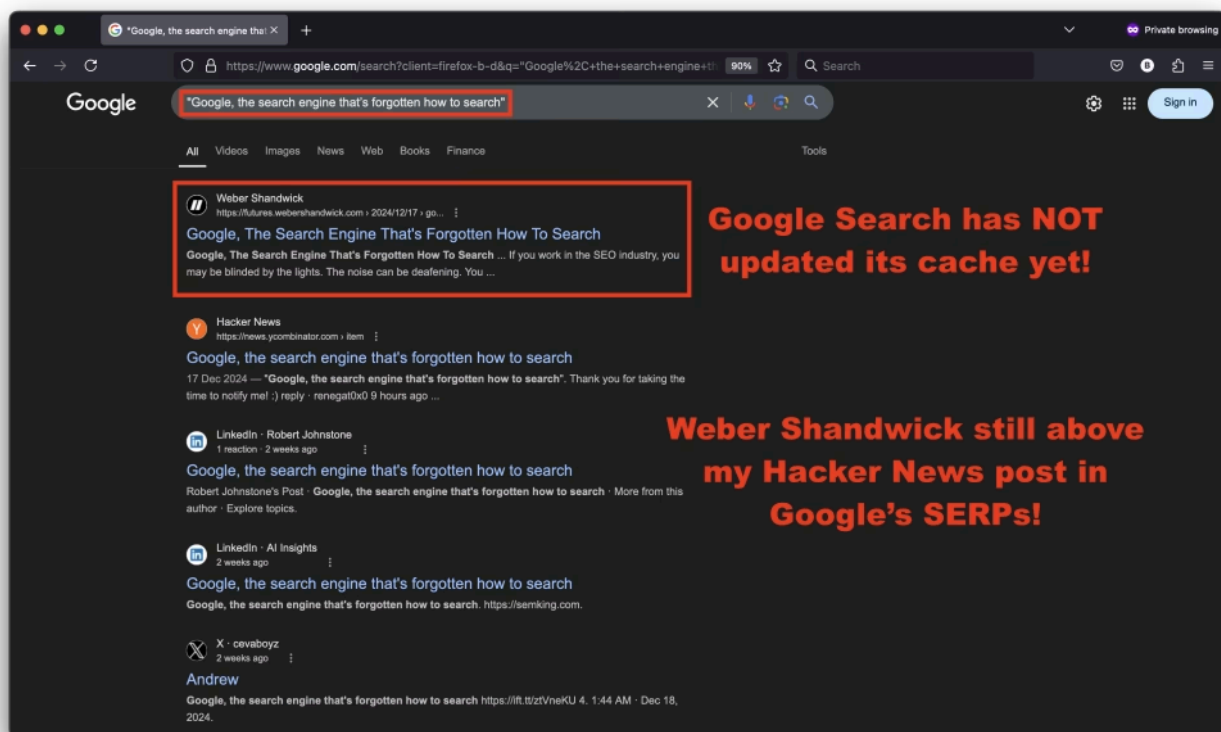
© [Elie Berreby](#) — This case study was first published on my personal blog [semking.com](#) on January 2, 2025 at 1PM UTC+0 = 8 AM Eastern Time (EST). If you read content this anywhere else, HELLO! 😊

UPDATE 1: January 3, 2025

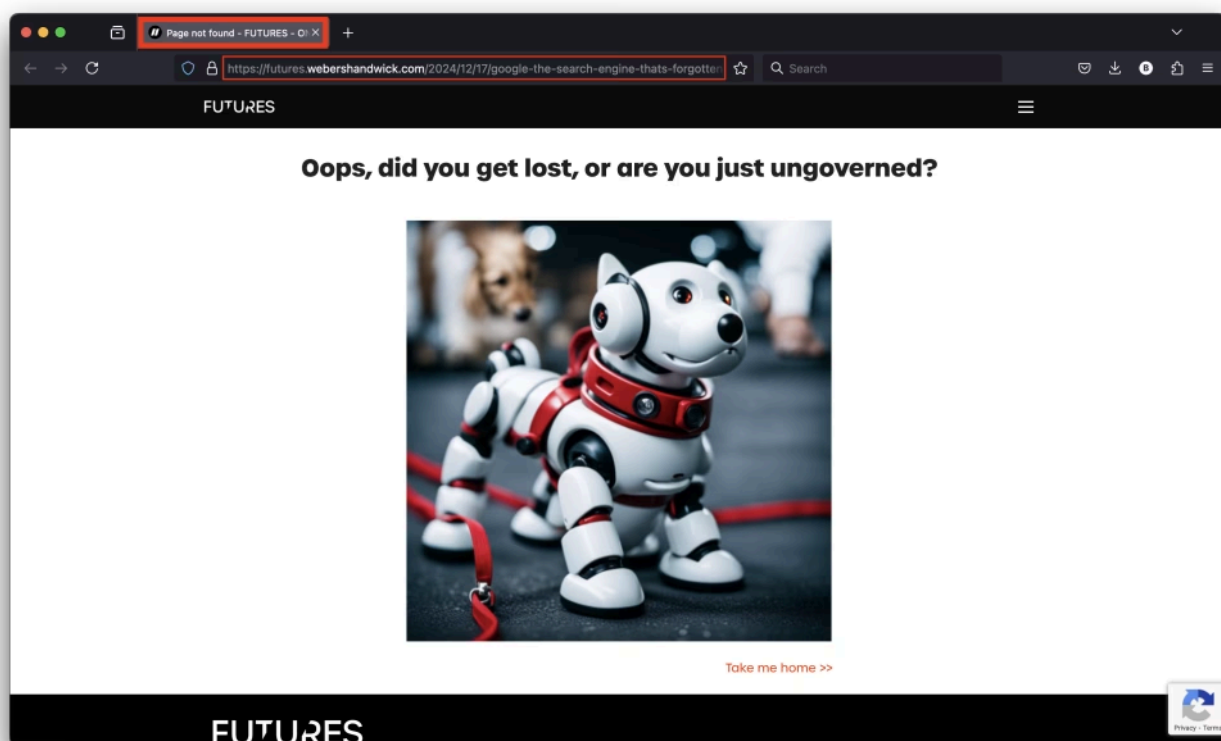
Here's what I see when I Google my previous article's exact title:

"Google, The Search Engine That's Forgotten How To Search"

For me (Europe), the Weber Shandwick subdomain still ranks at the top of Google's SERPs (above Hacker News), but only because Google Search has not updated its cache yet! Because I noticed something just happened.



Weber Shandwick has decided to delete their scraped version of my December 2024 article in the last hours!




I asked 3 of my core followers to check Google's SERPs from their respective locations. Bruno was the most reactive! Here's a 1 minute video from him: he's connecting from Portugal, with a Portuguese IP address and he uses the Sidekick web browser.

Thank you, Bruno! Do I believe it would have been better to correct the syndication issue, fix the authorship/attribution problem, and create a backlink to the original website? I'll let you guess. **IMPORTANT: read UPDATE 2 below!**


UPDATE 2: Weber Shandwick's reaction – January 3, 2025

I didn't tag the Weber Shandwick brand account anywhere and yet, they not only read this deep article, but also reacted in the comment section of one of my LinkedIn posts. Impressive!

**Weber Shandwick**40m
418,609 followers

Elie Berreby Thanks for bringing this to our attention - we discovered and are investigating an issue with the feed which made your content come through as incomplete. You'll see your link is now available through our site with proper attribution.

Celebrate · 🎉 3 | Reply

**Elie Berreby** **Author**5m
Senior SEO n00b

To the individual(s) managing the **Weber Shandwick** company account, you are doing a great job regarding brand monitoring! I didn't tag you on any social platform but you noticed my article.

Reputation management at its finest! Bravo! 🙌🙌

First, allow me to give credit where credit is due: the people managing this company account did a great job! Weber Shandwick manages their online reputation well! Congratulations to all the individuals working behind the scenes at Weber Shandwick. Let's not forget that behind brand accounts, large and small, there are anonymous professionals working hard to protect brands and companies!

3:07

53



futures.webershandwick.com



FUTURES



Home > Inspiration

Google, The Search Engine That's Forgotten How To Search



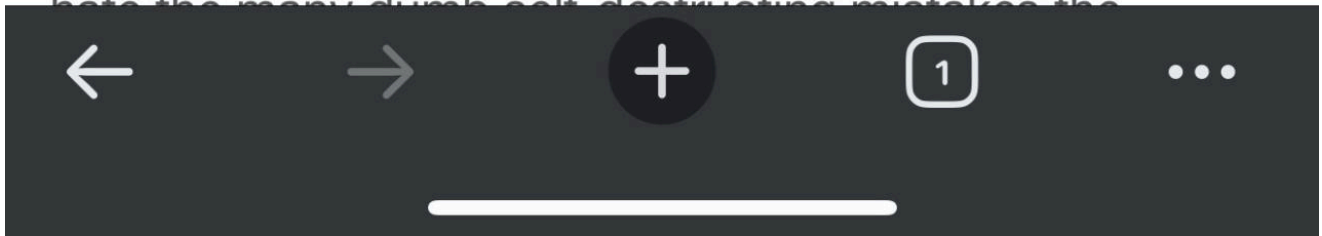
By **zapier** — 2 weeks Ago



If you work in the SEO industry, you may be blinded by the lights. The noise can be deafening. You may trust most of Google's official statements. Many in the search community blindly believe Google but I'm glad not everyone does.

I've been increasingly critical of Google Search in the past years but I'm not alone: many technically unsophisticated people using Google have also noticed the decreasing quality of organic results.

I'll show you why ordinary people are unhappy with Google Search, Google Ads and YouTube Search, using their own words. I'll explain where I found the less biased conversations. Senior people at Google have blocked me but I don't have a personal vendetta against Google. It is actually quite the opposite. I used to love Google. I just hate the many dumb self-destructing mistakes the



Secondly, speaking of companies, let's talk about Zapier because before Weber Shandwick's reaction, Kim discovered something interesting.

Yesterday, she was browsing the exact URL of my scraped article on Weber Shandwick's subdomain but not a desktop computer: on a mobile device.

It appears Weber Shandwick relied on Zapier to automate their feed. Kim shared this exact screenshot taken on her smartphone via a LinkedIn comment.

This suggests that the Weber Shandwick feed on their futures[.]webershandwick[.]com subdomain was automatically posted by a "zapier" account —probably the easiest name to track what was happening in the back-office!

I'll let you know if Weber Shandwick shares anything technical regarding what happened but this was interesting.

I want to thank Kim because I rarely use mobile devices for research and I probably would have never seen this on desktop 😊

Her original comment:



Kim Kuhlman, PhD · 1st

B2B Content Marketing Strategist | SEO | AI Expert | Data-driven busine...

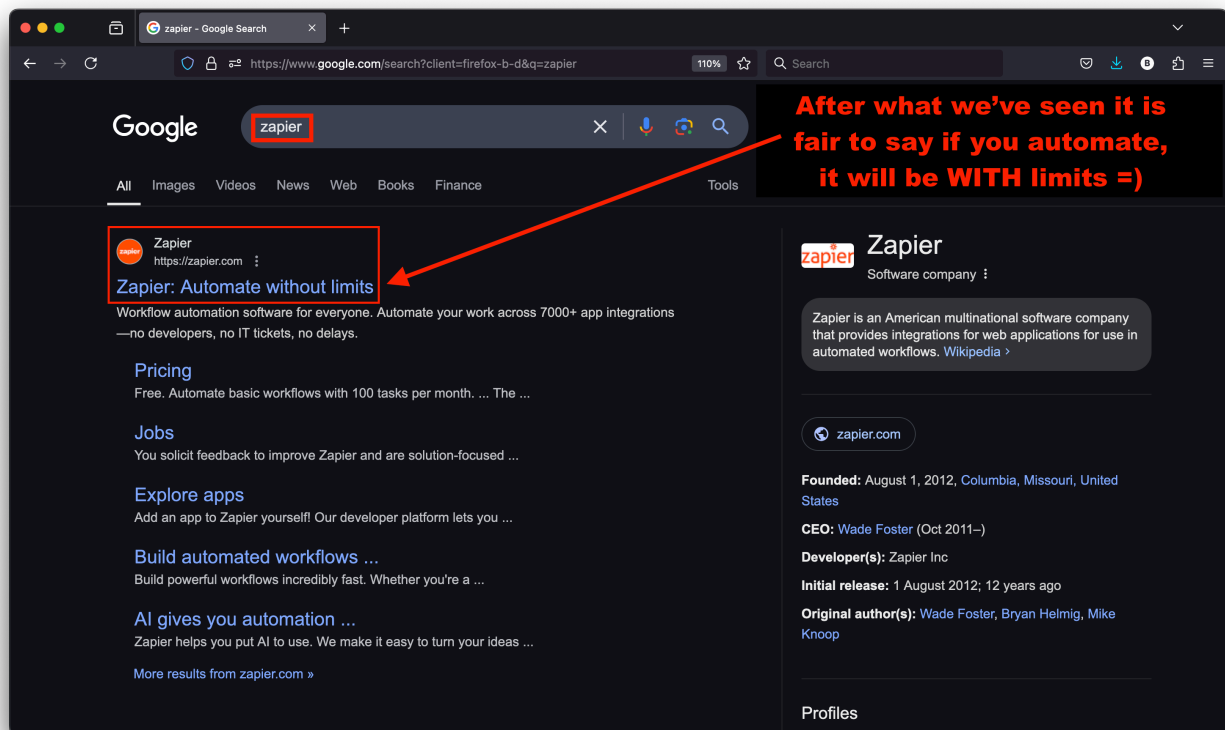
1d ...

Great post [Elie](#)! SO happy to see you're publishing on owned land first. 🥰
I'll be very interested in what you find going forward. I too use GSC to get content crawled.

A couple of things I found:

1) It looks like futures[.]ws is using **Zapier** to republish articles verbatim. (See screenshot). I saw this using G & B in incognito mode. Neither shows me that URL in search in a normal browser window. The Zapier integration is likely why you're seeing the "sloppy" scrape of HN.

I'm sorry, but I now have an intense and almost uncontrollable urge to roast Zapier... an urge that is only counterbalanced by the fact that if Zapier were indeed the invisible culprit in this whole experience, I would never have written such a comprehensive article without them! I believe a sober comment is enough: after what we've seen, isn't it fair to say that automating WITHOUT LIMITS can be risky? 😊

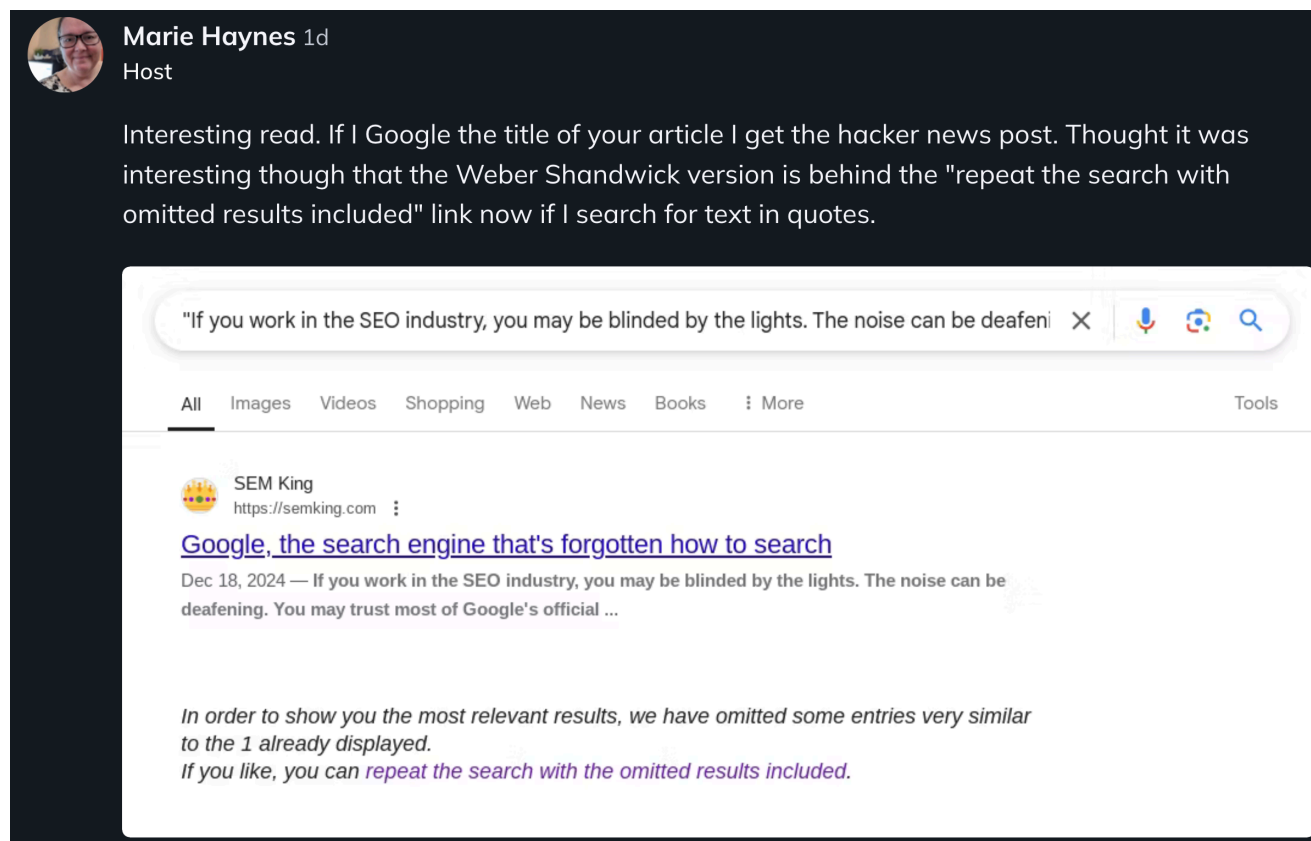


Google rankings after Weber Shandwick's traffic redirect

Marie Haynes read the first version of my article (before the updates). She shared this interesting observation yesterday, showing that my article URL was in the OMITTED search

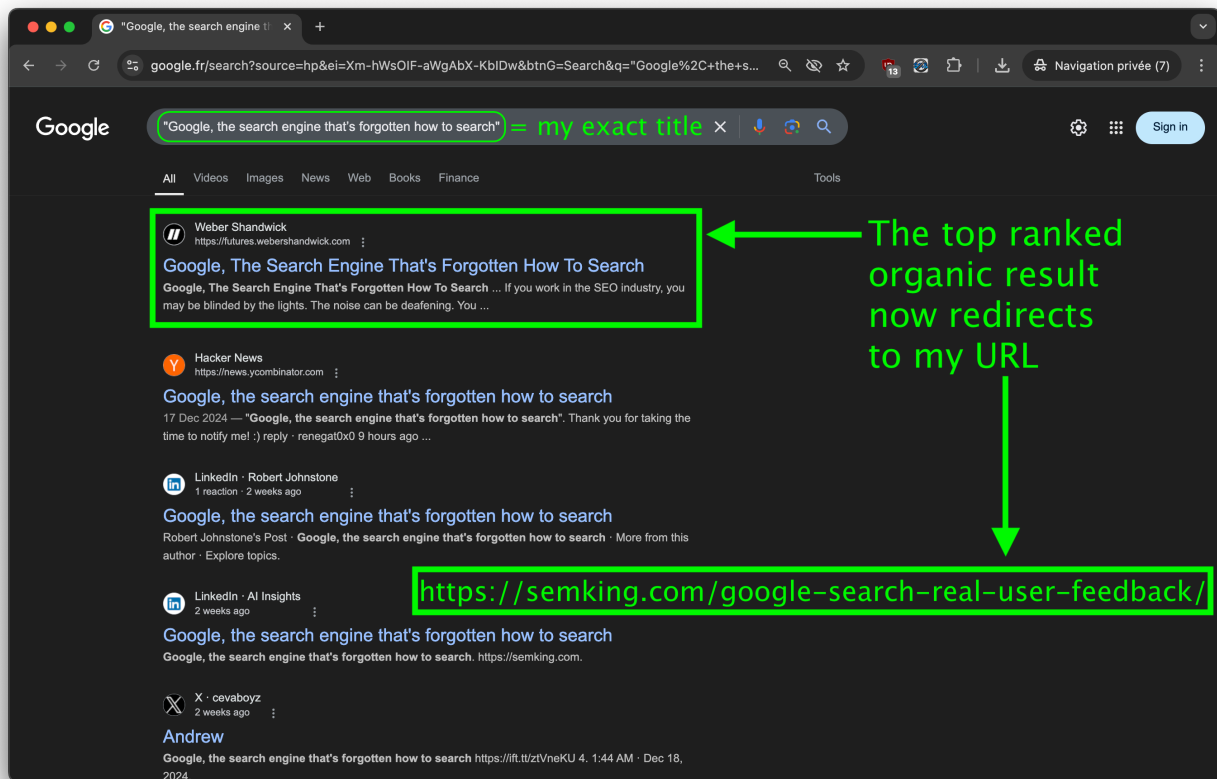
results!

That's shadow-banning 101: you exist but you are nearly invisible.

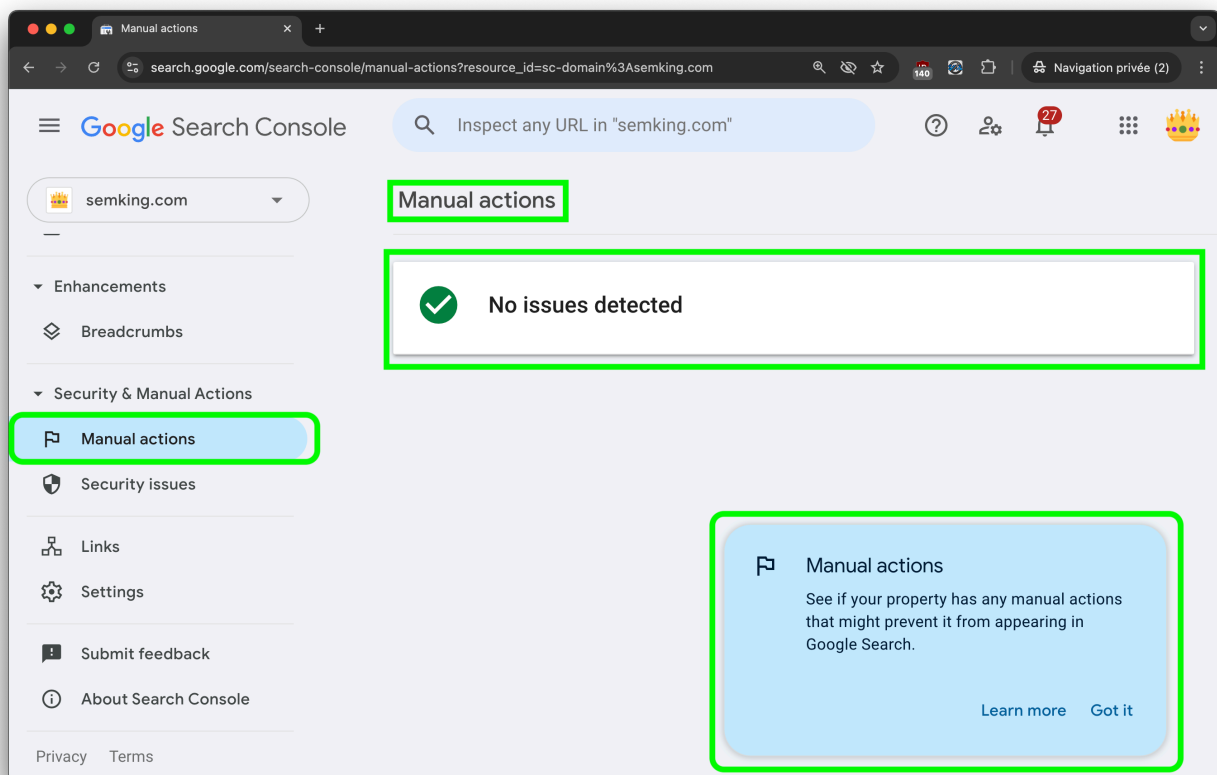


Weber Shandwick's decision to redirect traffic to my original URL is incredibly positive because Google's search engine result pages (SERPs) must now evolve. As you can imagine, I'll pay close attention in the coming days and weeks.

If you do not understand why Google's SERP must now evolve, let me explain in a simple way. The website currently ranking in first position on the first page for my exact article's title just decided to redirect its traffic to my article's exact URL.



In a few days at most, Google's SERPs should reflect this... unless they did shadow-ban me intentionally. I'll let you imagine the public outcry if it turns out that Google shadow-banned an article critical of Google Search. In that case, the conclusion would certainly not favor the search engine. Again, Streisand effect.



I never had any manual penalty in Google Search Console, but we know that Google people read my articles, including this one. Let's hope this was not intentional on Google's end. I would not have thought this was a credible possibility, but I now have doubts after seeing what they manually did to Forbes Advisor around September 2024.

- A good move would have been to work on an algorithm update to address all violations, at a global scale, of Google's site reputation abuse policy concerning affiliate practices.
- A bad move would have been a manual action that would not solve the global issue. Escaping a manual action is extremely easy. Any senior SEO knows what to do.

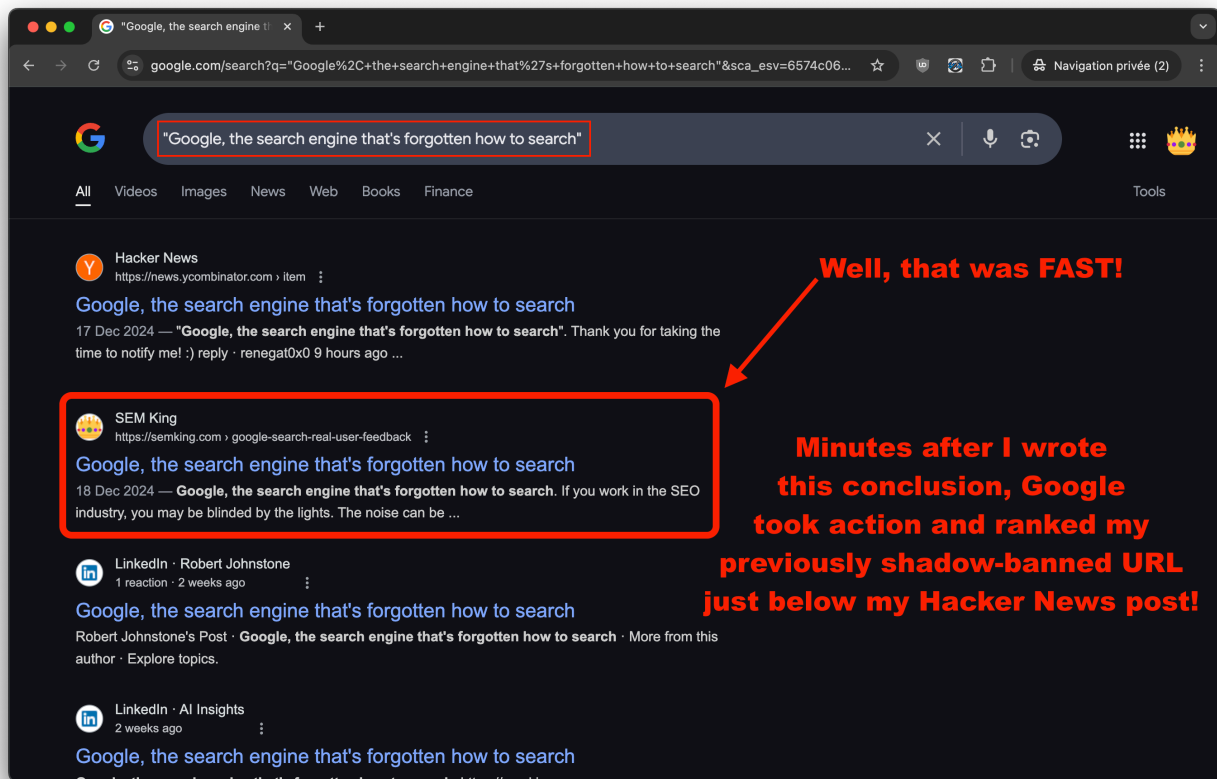
In my case, there's no manual action... and even if there was, I said I would opt for a passive approach, and I will —for a while. I'd like to see this whole situation resolve itself on its own. Now if in a few weeks, we see that Google still shadow-bans my critical article, everyone will reach an obvious conclusion!

UPDATE 3: Google is watching me (hey Google)!

UNREAL! Minutes after I wrote this updated conclusion and after more than 2 weeks of complete shadow-ban of my article's URL, Google took immediate action. Here are the NEW rankings for my exact title!

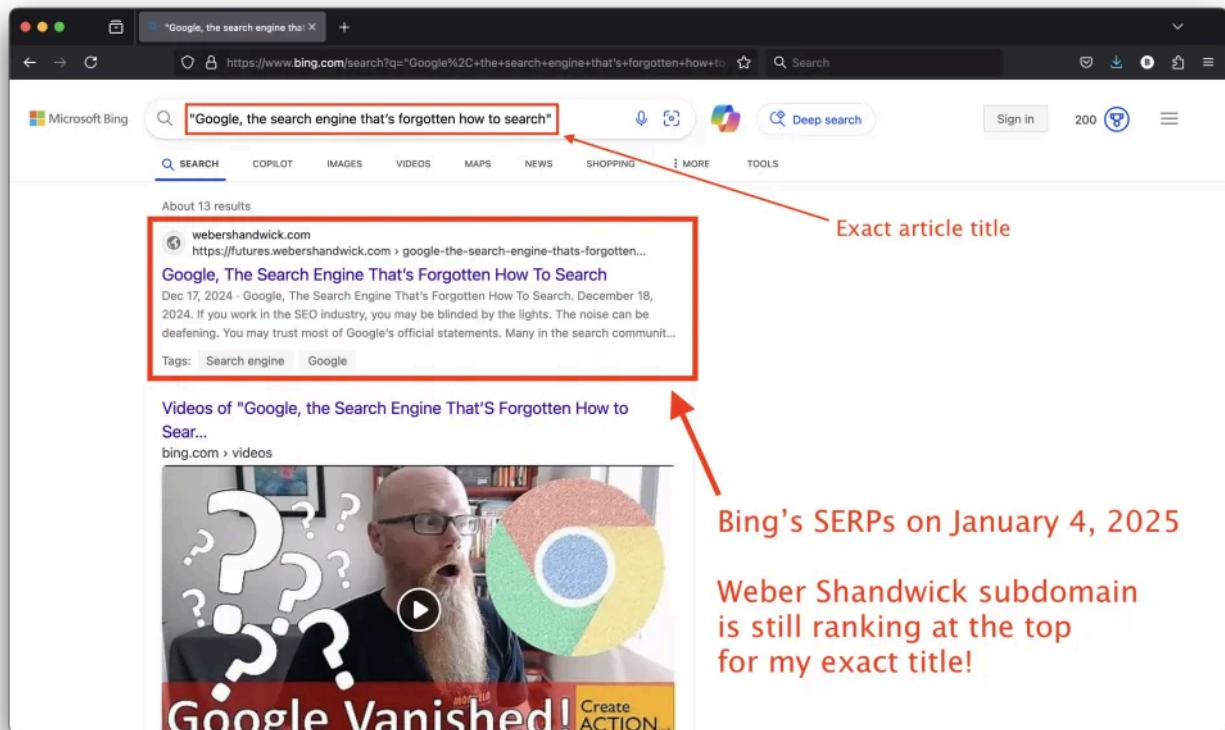
But now, I'd like to solve the final mystery: how can Google display my original article as published on 18 Dec 2024, and my subsequent Hacker News post as published on 17 Dec 2024?

This is unbelievable: I created both entries myself! First my article, which was indexed in Google, and then my Hacker News post. LATER, not BEFORE! Googlebot, you are so drunk!

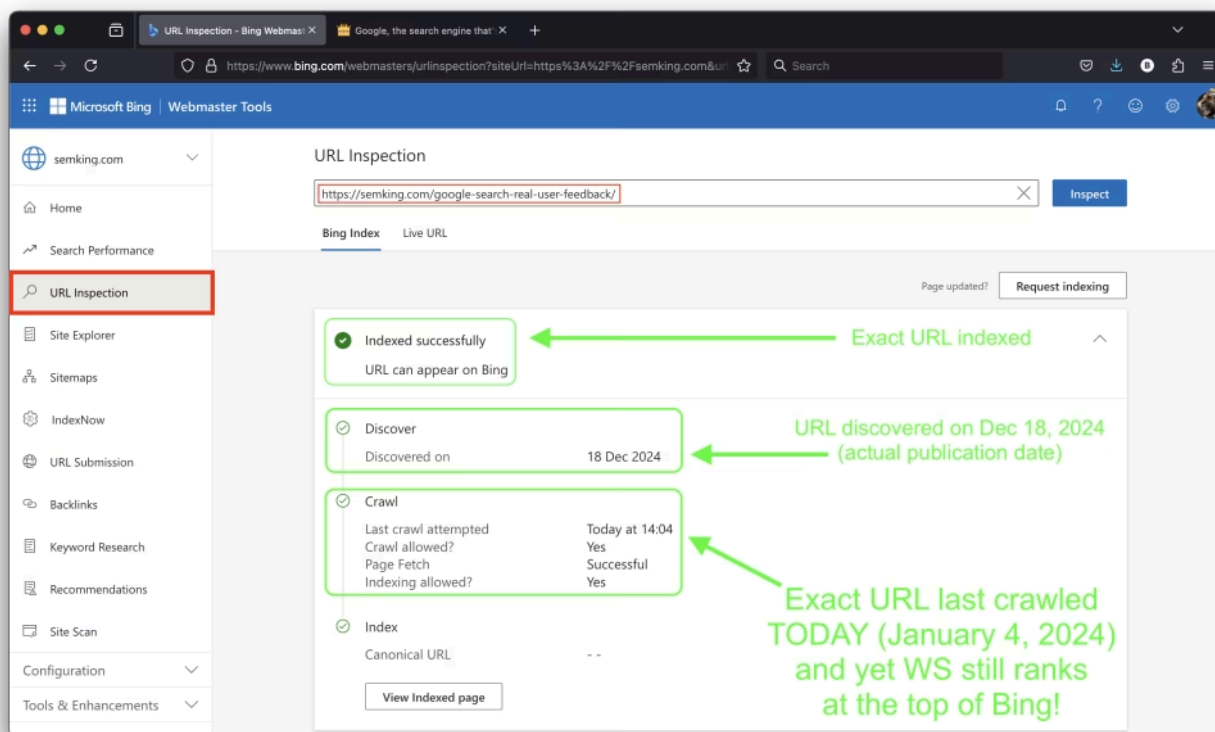


Update 4 (January 4, 2025): Bing's SERPs

Here's a screenshot of Bing's SERPs on January 4, 2025: Weber Shandwick is still at the top (first position, first page) via its subdomain for my exact title.



Could it be because Bingbot crawls content less often than Googlebot? Let's have a look at the URL inspection tool in Bing Webmaster Tools. Bingbot last crawled my exact URL about 2 hours ago!



I monitored, analyzed, and tracked everything following my initial Hacker News post. While performing log file analysis (Debian GNU/Linux server + Apache web server), I discovered something interesting regarding Bingbot's crawl budget for tiny websites like mine. I plan to share this soon, but to summarize, my discovery is unexpected and counter-intuitive!

Gratitude Overflow: Thanking the People who Shared my articles

I want to thank Barry Schwartz for including this article in his newsletter. This is the second time in a row in a few days that I had the honor of being featured on Search Engine Roundtable. I suggest you subscribe to keep your hand on the pulse of the Search industry!



Search Engine Roundtable :: TT: X

https://www.seroundtable.com


90%

Search

SEARCH ENGINE ROUNDTABLE

[Google](#) | [Bing](#) | [Algorithm Updates](#) | [Maps](#) | [Ads](#) | [Search](#) | [Subscribe](#) | [More](#)

January 2025 Google Webmaster Report



Subscribe!

THE PULSE OF THE SEARCH COMMUNITY


FOLLOW [f](#) [x](#) [t](#) [v](#) [r](#)

[Subscribe Options](#)

SEO

- How to Remove Your Site from AI Overviews in Google Search. Mr Jonathan Jones
- Search Engines Think I Plagiarized My Own Content! My Hacker News Case Study. SEM King**
- Understanding the December Google updates. Have you been impacted? If so, what do you do?. Marie Haynes
- What Is Google's Navboost Algorithm?. Semrush


Latest Articles

**Barry Schwartz** · 3:34 PM

ill include that in my newsletter tomorrow and subscribe to your RSS feed
good work!

Here is a recap of what happened in the search

Search Video Recaps





Bruno Costa · 3rd+
IT Advisor & Consultant
1d · Edited · 🌐

+ Follow ...

An exceptional article from **Elie Berreby** how domain authority is shaping the future of digital content. 🎯

👉 The direct link to the article can be found in the comments. 👉

Although extensive, it touches on a critical point that some have addressed, but not as Elie detailed, experimented, and documented: the growing dominance of big brands at the expense of the struggle of small content creators.

With AI changing the game and the issue of "intelligent paraphrasing", this article couldn't be more timely. Moreover, the rise of big brands has often been achieved through bad intentions and unethical practices, and this must be said without any filters.

The final infographic is particularly valuable, resuming the complete cycle and necessary checks for writing and syndicating content on high-authority sites.

A must-read for anyone creating online content in 2025! 💡

#SEO #DigitalContent #ContentStrategy #WebPublishing #digitalmarketing



Elie Berreby · You
Senior SEO n00b
2d · 🌐

I didn't write this contrarian indexing article as an SEO strategist. I wrote this as a tiny web publisher who creates content for various audiences. Few read

I also want to thank Kim for her comments and support. And Marie for her feedback.

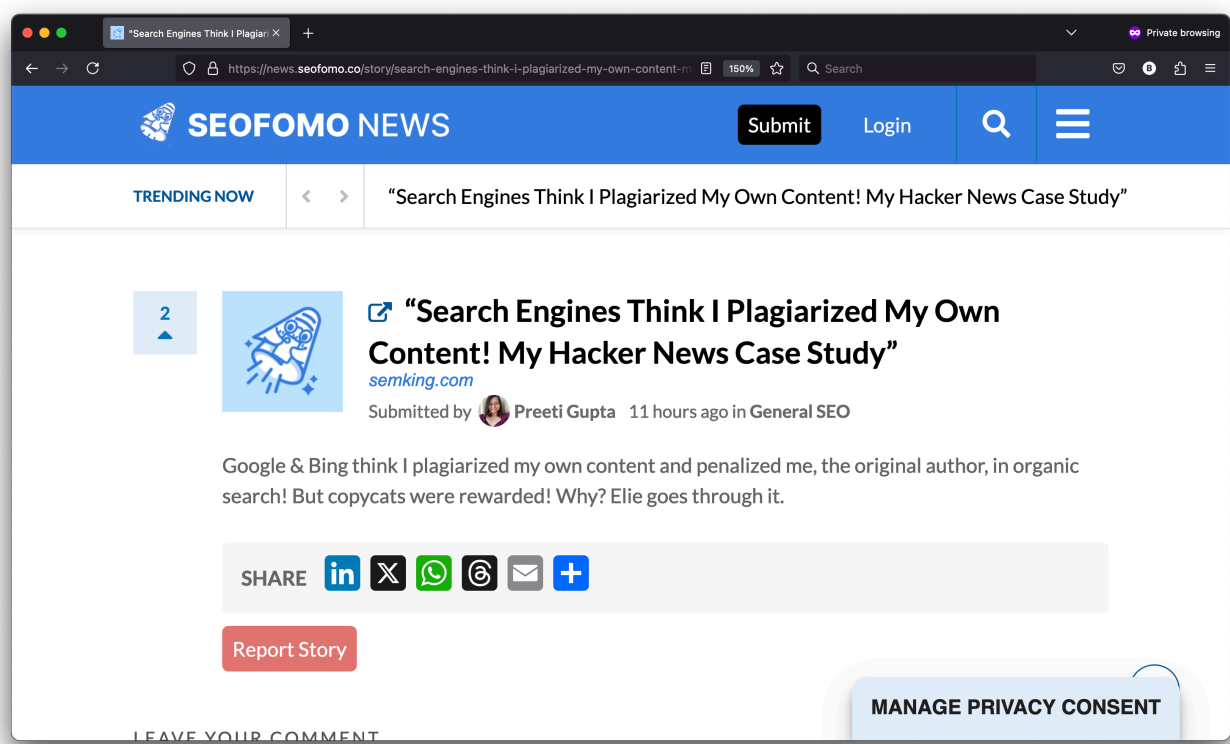
I want to thank Bruno for his short video and for sharing my article with his network: I appreciate quality interactions and I often repeat that I'd chose 50 creative followers over 50,000 bots.

Social proof is NOT an abstract concept, it is a reality. But the recognition we get can come from fake metrics.

Online, we should assume everything is fake. And we should be positively surprised when interactions are genuine and qualitative!

I also want to thank Preeti for secretly sharing my article on SEO FOMO News. She's one of my long-time followers, but I had no idea she shared this content.

I was analyzing the traffic sources when I noticed her post! I wanted to tell her I was grateful and to do so, I was “forced” to join the SEO FOMO community to leave a comment — it is still pending approval! 😊



A few hours before I finished writing the original version of this article (without the latest updates), a former strategist at Mozilla promoted my previous article (still shadow-banned at the time) about the quality decline in Google Search and user frustration. This led to some great backlinks and extremely attentive readers spending a lot of time reading this case study.

AS	Root Domain / Category	Backlinks	Country / IP	First Seen	Last Seen
95	mozilla.org	> 73	15.19	2d ago	Yesterday
New					

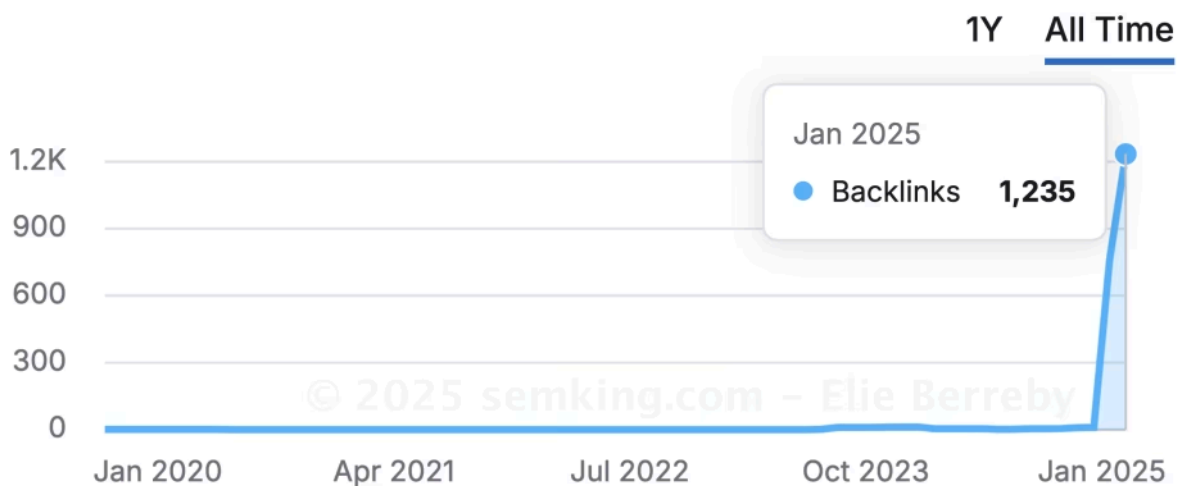
I’m not surprised. I’ve always admired the seriousness of the Mozilla community. I’m using almost every web browser for research and among those are Firefox and Firefox Focus, which I use for hours every day.

A couple of people I pay attention to began sharing my article on their websites. Here’s how the backlink profile of my website semking.com evolved in the first days of January 2025. A few new trusted domain names are now linking to my blog. This is not yet reflected in Google Search Console.

Referring Domains [semking.com](#)



Backlinks [semking.com](#)



Now, if you plan to share your viral or high-quality content with the world BUT want to avoid going through such a wild ride, here's a simple infographic I created to list the steps you should take before and after! 😊

STEPS FOR SHARING YOUR QUALITY CONTENT



© 2025 semking.com

Sign your texts & images

Multi-modal LLMs scrape and digest every format: sign your texts and watermark images!



Send strong quality signals

Instrumentalize tools such as Google Analytics to send as many qualitative engagement metrics as possible.

Think of others (Jon)

Jon's email showed me the importance of accessibility. Offer dark and light modes and add PDFs!



Search intent

Try to understand your audience's search intent (do the opposite of what I did).

Speak to copycats

Unlike me, don't be passive: talk to large websites that potentially hurt your visibility in organic search.



Digital PR or social media?

Before sharing on Hacker News, make sure you already exist in organic search, in the press or in social media posts (**Linkedin & X**).

Active Monitoring

Track every indicator: indexing for your keywords, SERPs evolution, social media and press mentions, user behavior...



© 2025 Elie Berreby – semking.com