

How to read the Indexing Report

31 MIN · YOUTUBE · [HTTPS://WWW.YOUTUBE.COM/WATCH?V=1-PCVLHPLWC](https://www.youtube.com/watch?v=1-PCVLHPLWC)

<https://www.youtube.com/watch?v=1-PcVLHp1wc>

SUMMARY

In this episode of "Search Off the Record," Mario Schmidt and John Mueller discuss common misconceptions about the indexing report in Google Search Console. They emphasize that the report should not be viewed as a checklist of issues to fix, but rather as a tool for identifying patterns and understanding site performance.

- The indexing report can show pages as non-indexed due to various reasons, not necessarily indicating a problem.*
- Users should focus on patterns and unexpected changes in the report rather than treating it as a static inventory to monitor.*
- Errors like 404s can be expected and do not always indicate a quality issue; they may simply reflect site changes.*
- Google's indexing decisions can be influenced by overall site quality, not just individual page content.*
- The "mark as fixed" feature is useful for notifying Google of significant issues that have been resolved.*
- Understanding canonical URLs is crucial, as Google may choose different versions based on external links.*
- The ratio of indexed to non-indexed pages does not inherently reflect site quality; many successful sites have a high number of non-indexed pages.*
- Monitoring trends in the indexing report can help identify systemic issues with hosting or site configuration.*

Bonjour Bonjour. And hello. My name is Mario Schmidt, and I welcome you all to a new episode of search off the record, the podcast where we the Google Search relations team, are taking you behind the scenes of Google Search and hopefully have some fun along the way. With me today is John Mueller, who's the mastermind of our team. Hi, John. Hi, Martin. So good to see you. Or hear you. Yes I have something I want to talk to you about and double check that I got it right.

OK I don't know if you remember, but recently, at a Search Center live event, there was a question about the indexing report, the coverage report in Search Console where someone was like, well, there's this page that I have and it's indexed, but the Search Console report shows it's not indexed, and why is it wrong. Was basically the question. But they were very friendly. They were friendly. You make them sound very angry, but I think they were friendly.

That's because I'm German, I think Nazis. Anyway, so I thought about this a little bit, and I think I know what happened here, and I recently saw another example of that with someone on Reddit who thought they screwed up because they have a bunch of non-indexed pages. Yeah, I think it's something that a lot of people, especially in the beginning when they start using Search Console, they run into this kind of thing where they look at the page indexing report and they see, oh, there are hundreds of pages that are not indexed.

This is an error, clearly on my side, something that I have to fix. And then they go off and try to do things to understand it or fix it. I mean, you even get emails about it and that kind of urges you to go into action. And I think the mechanism we have people to acknowledge that they understand what's happening. Is this fix, confirm fixed or mark is fixed or something. And I think while generally a great thing and I think in this case, it's triggering the wrong response or the wrong expectation I think.

Would you say that's maybe the case or I think maybe wow, maybe it depends. I have to do it. Depends otherwise this can't be a search podcast. But it's something where we try to use a similar UI across Search Console for a variety of different features, and that includes this marked as fixed button in a number of places. And in some cases, there are things that you can fix, if you have a bunch of 404 pages and you realize your server was set up wrong, you can tell us you fixed this issue and we'll try to fix it.

On the other hand, if something else went wrong in indexing, and it's more like our systems decided we weren't as curious as you might be then. It's not something that you necessarily fix yourself. It's not something purely technical. Yeah, right. I would generally agree. And I think I had a talk with someone from Search Console with Hillel from the Search Console team. Brilliant guy, lovely person, very helpful. And I think that opened my eyes to how we should probably talk about this and how people should probably look at this feature.

And it's slightly different from what I think people are looking or how they are looking at it now. OK tell me more. So all right, so he said, oh, I wish people would not look at it as a list of things to fix. That's what we have told people. And also not a static inventory that you need to monitor. Oh, is this page still in index. Is this page out of the index. What's happening here. Instead, he says, it's great to look for patterns and to look for things that are unexpected and then dig deeper into that.

It's an interesting way of looking at it, I think, because it's not like is this index does not index. It's a different thought. It's like, is the site in the index doing what I expect to do. And I think it can help you to spot things that haven't happened the way you intend them to happen. So for instance, if you're migrating and you're setting up redirects, then indexing report shows you to a steep rise in redirects page with a redirect.

It's not indexed because it's redirecting somewhere else. And if you look at it as an inventory, as a monitoring tool, only then you're like, what's happening. Why is there pages that are not indexed. Because they're not redirecting, when in reality that's exactly what you want them to do, because you just set it up. And on the other hand, a problem would be if that wasn't showing up. Well, a problem I say a problem, but it takes a while for search to see this change and actually process this change.

And that once it's reflected in the data search has understood this. As long as that hasn't happened in the index report, then OK, interesting. We have to wait a little longer because the processing hasn't completed yet. Yeah and I think that's a really powerful tool. And it also then doesn't matter as much that the data is not always real time. That's the other thing. People publish a new section or a new category on their website and then they're like they're not indexed.

And in reality, it might just be that we have a data delay or it might take a few hours longer. We sometimes do have longer delays, but we normally communicate these in Search Console. We right. We have these little notifications that we send out. But yeah. So I think that's really, really handy. The other thing is that sometimes you see changes that you haven't triggered. I mean with a redirect or if you remove a whole category of stuff on your website, then you expect to see a steep rise in 400 fours because you removed stuff from your website.

And again, if that hasn't happened, then that tells you something has gone wrong. Maybe or something hasn't happened yet or hasn't been processed yet. Yeah I also think especially with 400 fours, that's something that sometimes throws people off. Yeah because they're labeled as an error. And it's almost like your website is returning an error. And then a lot of people see that and say, well, I have to fix this error. Yeah and I have to make sure that Google doesn't see a 404, because if Google thinks my website has errors, surely Google will think my website is not as useful.

Yeah, it's not good. We need to fix this. It's an error. You want to fix an error. But it's an error that is expected. It's actually a good thing. The problem is that. And then I get the question like so. But why does Search Console show it as an error. Because it is an error. It's just an expected one. Yeah, I think that part is always tricky for people in the beginning and especially tricky if you have, I think, management that doesn't understand it so much.

And management tells you it's like you should make sure Search Console is always green. Everything is OK. And then people might be worried that it's like, oh, I have this error report and this number is not going down. And I was like, what. What do I need to do to make this error number go down so that my boss thinks I'm doing a good job. And actually, if you're removing things from the web or if these pages never existed.

Yeah, they're supposed to return an error. That's actually the right thing. Yeah, absolutely. And the other thing is, so it can't judge for if this is a problem or not have to do the judging yourself. And the other thing is sometimes it's changes that you haven't caused like page with a different canonical. That's something that makes sense to look into, but it's not something that should alarm you, because that just means that the report, as is now reporting for different URLs.

And I know that the numbers then sometimes look like it's going down, but then that usually means it's going up somewhere else. And I find that useful because it tells you how Google thinks about your site and your URLs. And if you have I don't example.com, it might be that external people are all linking to a different version of that, just example.com and maybe Google thinks, hey what.

This is shorter and everyone uses this. So maybe we use that as the canonical. And that can help you change your canonicals to something that everyone else would prefer anyway. And I think that's a cleaner way of doing things, but it's not a problem that you need to fix. Yeah, I think fundamentally the question of whether you should change your canonical to what Google thinks your canonical is or not, that's up to you. That's a very different question.

But the aspect of Google picking another URL within your website as a canonical, that happens all the time. And that's normal. And it's not something that you need to worry about because it's just it's still in the index. People can still find it. Yeah, exactly. Yeah that's something by the way, that is a little bit easier if you have a domain property set up in Search Console. Yeah, that's true because then you don't have to worry about.

And whether you should say W or dub or whatever, all of these things are a little bit easier. And basically all of that is shown in the performance report. That's true. I do think it is useful to see the site from bird's eye view. And I really, really enjoy the report for that. So I wouldn't use it to look specifically after like, oh, this page needs to be in the index and this page isn't in the index.

And what do we do now. But more like are we seeing some pattern moves and why are we seeing them. And for that it's great because you can see oh, pick the different canonical for a lot of pages. You can click on that get a bunch of examples, and then you can analyze one or two of these examples and then basically get a feeling for OK. So that's happening on the page. And sometimes you can basically just conclude well everything is fine.

There's a shift of something. But the shift was to be expected. Or at least the shift isn't malicious. So I think the way that Hillel presented this makes a lot of sense to me, because it's more like a don't think of it as inventory. And we have a similar problem with the site query as well. Yeah where people are like it's not showing up for the site query, so it must not be in the index. But Google Search Console says it's in the index.

So who's right. Yeah I think site query is an artificial type of query. So yeah, that's something where partially you can use it to double check if a URL is indexed. But Search Console is really the source of truth there because sometimes we just don't show things inside queries. And sometimes we do show things inside queries that are not actually indexed like that. For example, if you do a move from one domain to another and you do the move properly with redirects and all of that, then basically, if you ask Google about the old URL, the old domain with a site query, Google will tell you it's like, oh, yeah, I know about this.

And it's like, here's a bunch of URLs. And this will happen maybe a year or two or even three afterwards after you do the move. And if you look at purely the site query for that might conclude that, oh the site move actually didn't work. Google is still indexing all of these old URLs. And sometimes when you look at those results, you'll see in the title maybe the new domain name or something like that. So you can tell Google saw the move, but it still knows about those old URLs.

And if you explicitly ask it about something, then Google will be like, sure, I can give you what you're asking for. It's like, I'm not saying that this is the current source of truth, but it's like you're asking for this site and here it is. And it might be if you're a normal user, you might have remembered the old domain. You're like, what happened to this old website that I really liked. And if you look it up on Google, Google will tell you it's like, here's a content that was there, and if you click on it, you'll of course, be redirected to the new site.

Does that sometimes also happen if you have hreflang setup and have different language versions. Probably because the way hreflang works is the individual versions have to be indexed, but then when they're served in

the search results, they're swapped out against the local version. And I don't for example, what happens with a site query. If you do, I don't site colon something and you're in Germany, maybe you see the German version, maybe you see the Swiss version, I don't know.

Interesting someone should try it out. Please try it out and write it in the comments. We are curious to see how that happens. That would be interesting. Yeah, I think maybe taking a step back to some of the errors that you mentioned. One of the things that I think is worth watching out for is whether or not there are any systemic issues with regards to your hosting setup, and this is something that you do see in the page indexing report as well.

And it's the kind of thing where you look at it and you're like, well, this is unexpected. Why are so many URLs suddenly considered 404 or 403 or I don't dropped for other reasons or even canonicalize to some other page. And sometimes we've seen CDNs, for example, or hosting providers. They have some kind of bot protection, and maybe they turn it on when there's a lot of crawling happening. And depending on how this bot protection is set up, it can happen that suddenly your CDN or your website host or serves 404 or 4, 10 or 403 or something instead of something reasonable, at least from what I would consider reasonable, something like a 503A that basically says, oh, I can't deal with you right now, come back later.

Maybe if they think Googlebot is a malicious bot as well. Maybe a 404 or something makes sense, but that's something I think worth watching out for. Every now and then taking a look at that report to see has there been a significant change in the number of crawlers or in the number of pages that are dropped from indexing, and if so, what are some examples that I can follow up on. The other one, which I think is a bit harder for sites to diagnose, is when the hosting provider shows something like an interstitial, something like are you a bot or not.

And it serves that with a 200 response code and you can guess what happens. It's like we'll try to index that on the one hand, which will result in your page's content kind of disappearing and being replaced with this. Am I are you a bot page. But what also happens is because this page is shared across a lot of other pages on your site and maybe other pages across other websites. It can happen that we pick a canonical of someone else who has the same error page, and that's really challenging to debug, because you basically have to look at those pages and look at the canonical or look at the page that Google chose as canonical when you look that up.

And sometimes you have to almost guess a little bit of what would happen when Google accesses this page. Because as a user, you might not see that interstitial at all. But if you see a lot of pages on your site being duplicated eliminated in favor of some random other page, oftentimes that's a sign that you have some kind of a soft error on your page, some kind of soft block that tries to block crawlers and bots and basically just ruins your SEO.

So that's something where of course, that's the kind of thing where I would say you should go off and try to get that fixed with your provider or your providers if you have a CDN and hosting. Figure out where it's coming from. And then using something marked as fixed in that report makes a lot of sense, because then you're really saying there was a significant issue on my website, like you found all these pages and you canonicalize them

somewhere else, or you called them 404, but actually they exist.

How dare you take away my pages. And then we will try a sample of those. So the way the marked as fixed works is we try a sample of the pages that you're basically telling us are fixed. And if we see that they're actually fixed, then in most cases, we will trigger a faster recrawl of the other pages. So it's not so much that we wait and oh, well, we'll see if this is actually working better, but we'll try to recall that a little bit faster.

That makes sense. And again, this is about patterns. So if you see a bunch of arrows then you can figure out. So this is all the products because x, yz or this is all the sides. So this is not just a specific template that is misbehaving. And it helps you troubleshoot a little bit faster. Because if it's just one specific template where pages are experiencing some specific problem, then it's probably something within your hands, so to speak.

But if it's the entire site or an entire part of the site that is behind the CDN, then it's probably something to do with the CDN. And again, these patterns are worthwhile. The same with the trend lines. I think we haven't really talked about the trend lines in the report yet. So not only do you get the reason for a problem, but you also get this line because sometimes a server has a hiccup. There's some sort of problem.

We are seeing some 500 something error, which can or cannot be a problem. That depends. If you're seeing a little blip, then maybe look into the server logs, what happened here. And then someone tells you, oh yeah, we upgraded the host. We moved between different computers in the data center. Whatever and then there's not much you can do about it. And it's kind of fine because it's going to work itself out. But if it's a steep line going up.

Maybe look into. Yeah, yeah. On the other hand, you might also see that. And it's like a horizontal line. Then it's a few blips and it doesn't. Yeah, yeah. I think it's important that people realize that computers don't always work. And for the most part, the internet is built in a way that is supposed to be resilient against that and just retry. Yeah so when we look at the crawl stats of a website, every now and then you'll see things like oh, I don't a handful of requests failed or a handful of DNS lookups failed.

And of course, we want to report that because maybe you care about those handful. But for most sites it's something where it's like, oh, this didn't work this one time when Google tried it, but the next time it worked and everything kind of worked itself out. And it's not going to be the case that if Google runs across an error and it just exists for a couple of minutes or whatever, that it's going to cause any issues across the site, even for the most part, when we run across errors and they exist maybe for a day, which is a really long outage in internet time.

That's something that we try to look over and say like, oh, maybe it's something temporary. We'll just like, keep things stable and come back and check again. But if it's longer than a day, then, of course, our systems are going to react to that. Definitely and you want to have a look at it. And also if you add or change your site or if your site is very new, then you can actually also use this report to see a little bit how your site goes through the different stages, because at some point, you're going to see pages in Discovered currently not indexed.

Which tells you we know they exist, but we haven't actually visited them. And if we haven't visited them, we can't put them in the index versus. Crawled currently not indexed, which means we visited them and we didn't put them in the index. And that can have all sorts of different reasons. Would you say that is often or only sometimes a sign of a quality issue. So it's definitely the case if our systems are seriously worried about the quality of a website, that they will reduce the number of pages that they index.

Because if we have strong concerns about the overall quality, then it doesn't make much sense for our systems to spend a lot of time on the website. So we'll probably crawl a lot less, we'll index a lot less, and then you'll see things like crawled, not indexed or discovered, not indexed, which from our point of view is basically our system saying, we know about this, and once we're happy we looked at it, once we're happy, we will take another look and see if we can index it.

It's not so much that I would say you should take these situations and try to fix them. From a technical point of view, it's not that you need to fix this technical issue that Google is not indexing this page at the moment, but rather you almost need to when you recognize a bigger pattern like this, that Google is not indexing a lot of your pages, and there's no technical reason, you almost need to take a step back and think about the quality overall.

And thinking about quality is really challenging because a lot of times, it's your website and it's your baby. And of course, it's the best baby ever. But taking a step back and trying to look at it with the eyes of someone who is not directly involved with your website. Sometimes that opens up some ideas for areas where you can improve, where maybe if most of your website is AI generated and it worked for a while, it might be that people look at this AI generated site and they're like, well, I can tell this is AI generated.

There's nothing unique or valuable that is available here for me. That's not to say that all AI generated content is bad, but sometimes you just run across websites where you're like, anyone could have written this. This tells me nothing. Yeah, that's true. And I think what makes this difficult is not only the fact that obviously the way you wrote it is the way you thought is best, and that's why you think it's high quality, of course.

So that's really, really hard to step out of your own perspective. But sometimes it's also there's so much other stuff that is just as good. So why would we add it to the index. And then that can tell you, like maybe this content isn't as valuable as I thought it is because other people are covering the same thing. And then what's the value of this version of it being in the index. Yeah that's true. I feel we could have a whole podcast about quality.

I think maybe one other thing that is worth mentioning with regards to quality is it's not just the text. So a lot of times people will say, well, my text is unique or my articles are good and they're packaged in a page that is terrible to access where anyone who, when they try to load it like their computer fan spins up and they're like, oh my gosh, I have to run away to make sure my computer doesn't explode.

So maybe that's an extreme case, but you've all seen these pages where basically the text is there, but it's almost hidden away, hidden behind ads, hidden behind interstitials, hidden behind other things that are moving and coming and going, maybe hidden below a bunch of filler content, which we sometimes see, for example, with

recipes where there's this really long story on top that maybe most people don't really care about. And then the recipe comes. These are all the kind of things where the overall quality is much more than just that piece of text that you say, this is my main content.

This is what Google should be counting for my site. And from our point of view, we almost have to take into account the full experience on a page, because that's what users see. It's not that users go to a web page and turn on some magic mode that just pulls out the text, but rather they have the full experience of this website with all of the 3D, 4D animations and everything. I agree very much. Agree oh my God.

The other thing is, there's no such thing as a ratio between index and non-indexed pages. I mean, there is, but it doesn't matter. I've seen so many websites that have a million non-indexed pages and half a million that are in the index and they're doing perfectly fine. So I think that's the other thing. A lot of people are seeing this and they're like, Google probably thinks my website is low quality because only I don't 20% of my pages are indexed.

No, that's not something that you need to worry about. That's not something that tells you like this is good or this is bad. Lots of it depends. Lots of it depends. I think that's something that people don't normally know. Yeah when I take a look at the Search Console for our developer documentation, for example, I think it's something like 5% of the pages are indexed. Wow which when you look at that report, you're like, oh my gosh, something must be seriously wrong here.

But when you look at the details, you see it's like, oh, a lot of them are no indexed. There's a bunch of 404 like canonical things the important content is all indexed. We see that important content also in the performance report. So all of that is OK. But if you just look at the page index of your report, you're like, whoa, this looks like something is seriously broken. I wonder how this looked when we did the big migration from the old structure to the new structure that we did a few years ago.

That must have looked really chaotic. Yeah, I imagine some of that, but also probably a lot of code samples and probably I don't from the other different product areas that are hosted on the same site. Maybe they have different setups where they say, a lot of this content should be no indexed maybe. I don't know if you have APIs with older versions. Maybe you say all the older versions are no indexed, which is a decision that you can make and is fine from our point of view.

It's just in the reports you'll see that reflected, or you usually will see that because we look at these pages and initially go and say, well, maybe we should index this and then well, maybe not. Maybe interesting. But I think we're good. Yeah Anyway, I think that's summing it up roughly what I wanted to bring out there. So don't think of it as an inventory that you need to fix. Don't think that everything on the website needs to be indexed.

Don't think that the ratio between index and non-indexed is some sort of quality measure, or some sort of score that you need to look at. I think as long as everything you care for is indexed with some URL. And again, domain properties make this easier because you can see if canonicals just change and don't have to worry like,

oh, this dropped out because of canonical. And then you have to figure out, OK, so the new canonical that's fine.

Yeah so look at it as a detection tool for changes or patterns in the site that you expected or didn't expect. So if I make a change, it should reflect in this report in the pages there as well as if I didn't make a change and something happened. What happened. And is that a tendency. Is that a trend. Is that something new. Is that something that always has I've just seen for our devsite, actually, we have a lot of 500 errors, but I mean a lot.

We have some 500 errors, but it's stable. So if this hasn't been a problem yet, then it hasn't been a problem like a month ago. And it will not be a problem right now because it's the same amount. Some requests will probably fail for whatever reason. Maybe solar wind, who knows. Yeah, I think I looked into those for the developer site at some point, and some of those are things like API requests where the documentation links to some API requests, and then the developer site server says it's like, no, it's like that is a bad URL, and I'm not giving you 404 because I understand what you're trying to do, but you're giving me a bad request.

Oh, I see. Interesting so that's where that comes from. And that I think especially for things like technical documentation and website technical documentation, I don't know how you would call that where basically you're describing URLs on your website for technical reasons and you're giving examples. Then you see a lot of different kind of errors. Whereas if you have a blog about recipes, then you don't see a lot of bad API requests because you're not writing about any APIs.

I see. Interesting yeah, I wouldn't know how I would call these requests either. Anyway, so I think we looked at the indexing report quite thoroughly, and I think it makes a lot of sense to treat it differently than some people are treating it today. And I hope that we took away some scare out of this report and made it more useful for more people. So thank you all very, very much for listening. And thank you, John, for joining me and discussing indexing report with me.

Of course. Happy to be here. Thank you Martin. Thank you. So we hope you had fun because we know we did. And please leave us and subscribe and comment. If you are seeing different language versions in the site query. And if you want to hear more from us, definitely check out our future episodes. Maybe check out our earlier episodes if you haven't listened to them yet. And cheers and bye bye everybody. Bye we've been having fun with these podcast episodes.

I hope you, the listener, have found them both entertaining and insightful too. Feel free to drop us a note on LinkedIn or chat with us at one of our next events we go to. If you have any thoughts, let us know. And of course, do not forget to and subscribe. Thank you so much for listening and goodbye.