

Search Off the Record - 43rd episode

[00:00:00] 🎵 [music] 🎵

[00:00:10] **John Mueller:** [00:00:10] Hello and welcome to another episode of Search Off the Record, a podcast coming to you from the Google Search team, discussing all things Search and maybe having some fun along the way.

[00:00:23] My name is John, and I'm joined today by Lizzi and Gary from the Search Relations team, of which I'm also a part of. So how are you coping with the Swiss summer so far Lizzi?

[00:00:35] **Lizzi Sassman:** [00:00:35] I am not doing well. I don't think that I know how to cool the apartment appropriately. It just seems to get very hot, even though I am closing the blinds. And I don't know if there's a trick to it or I just picked the wrong apartment or what is happening. I don't know if you also have the same problem.

[00:00:54] **John Mueller:** [00:00:54] It's the same here. I don't know, in Switzerland nobody has air conditioners or like, almost nobody. So you have to cope with it. I think this might be one of the reasons why people in Europe just go on vacation for several months in the summer.

[00:01:09] **Lizzi Sassman:** [00:01:09] It's not hot in Europe?

[00:01:11] **John Mueller:** [00:01:11] Maybe you have to go to some place where there are glaciers like this, like live in a ice cage.

[00:01:17] **Lizzi Sassman:** [00:01:17] Gary is making a face.

[00:01:19] **John Mueller:** [00:01:19] How are you doing Gary?

[00:01:20] **Gary Illyes:** [00:01:20] I actually love it. Hence, the face. Also in my apartment, it's like 23 degrees Celsius, so it's perfectly fine.

[00:01:27] **Lizzi Sassman:** [00:01:27] I thought you were saying it was humid.

[00:01:29] **Gary Illyes:** [00:01:29] Outside it's humid. Not so humid anymore. I mean, inside it's not humid anymore. So that's great, and I love the heat.

[00:01:36] **Lizzi Sassman:** [00:01:36] So you are having a great time in the heatwave.

[00:01:39] **Gary Illyes:** [00:01:39] I'm having a great time. It's one of the reasons why I go to Singapore to be in humid, extremely hot environment. Well, used to go to Singapore. And now, apparently, the Singapore weather just came here, and that's very, very nice. I appreciate it.

[00:01:53] **Lizzi Sassman:** [00:01:53] So ideal Gary climate.

[00:01:55] **Gary Illyes:** [00:01:55] Correct, let's move to Singapore.

[00:01:57] **John Mueller:** [00:01:57] I usually enjoy the weather in Singapore, because then I feel a lot more happy going back to Switzerland where it's like, "Oh, it's only so many degrees here. This is nothing."
[laughter]

[00:02:09] **Lizzi Sassman:** [00:02:09] I can tolerate it. But they have lots of AC there. So, could you just avoid the heat if you wanted to?

[00:02:20] **Gary Illyes:** [00:02:20] Well, you do have to go outside sometimes, like for a commute, for example, you have to go outside.

[00:02:26] **John Mueller:** [00:02:26] So, talking about blocking. Last time, we talked about blocking sites in Search, kind of from an owner point of view. And now it's our turn to ask Gary how that actually works in reality from a search engine point of view.

[00:02:45] **Gary Illyes:** [00:02:45] What?

[00:02:46] **John Mueller:** [00:02:46] Yeah.

[00:02:47] **Gary Illyes:** [00:02:47] I was not-- I-- What?

[00:02:50] **John Mueller:** [00:02:50] Yeah, tell us how it really works, Gary. So I know you have like the crawling, indexing and serving side of things like those three big buckets. Are they buckets? I don't know.

[00:03:04] **Lizzi Sassman:** [00:03:04] Bins.

[00:03:05] **John Mueller:** [00:03:05] Bins, systems. The engineers definitely want to not have them be called buckets, I guess, because that sounds kind of awkward.

[00:03:15] **Gary Illyes:** [00:03:15] I think bucket is a good name for them.

[00:03:17] **Lizzi Sassman:** [00:03:17] I like it. Yeah, it's fun to say bucket, you now.

[00:03:20] **John Mueller:** [00:03:20] I work on the crawling bucket.

[00:03:22] **Lizzi Sassman:** [00:03:22] Yeah, the bucket crawl, the bucket team, bucket of crawling. It's a bucket of fun.

[00:03:27] **Gary Illyes:** [00:03:27] OK.

[00:03:28] **Lizzi Sassman:** [00:03:28] [laughter] Too far.

[00:03:31] **John Mueller:** [00:03:31] So looking at those three buckets systems, how does the blocking side of things kind of flow in there? Is it just like a flag that someone sets and then everything gets processed normally? Or how does that actually work?

[00:03:47] **Gary Illyes:** [00:03:47] As you said, we have those three buckets, and each of the things that we talked about in the last episode map to one of those buckets.

[00:03:55] I think we talked about robots.txt and then we talked about some meta tags, specifically robots meta tags. And then we talked about maybe about the removals tool a little. And each of these things actually map to something within Search, like one of the bigger systems that you mentioned: crawling, indexing and surfing.

[00:04:14] And I think that's why removals, in general, are so chaotic when we talk about them, because we have all these systems that have their own way to control what they ingest and how they process things.

[00:04:29] **Lizzi Sassman:** [00:04:29] Right, so certain things, certain methods can apply to various buckets or stages of Search. So you need to know how the systems work in order to know which method to pick. Would that be a right way to think about it?

[00:04:45] **John Mueller:** [00:04:45] I think so. I don't think it's necessary, but if you are dealing with removals often, then you probably want to understand those systems. Because if you are-- you can combine certain things in ways that will not work.

[00:05:02] **Lizzi Sassman:** [00:05:02] Mmm.

[00:05:03] **Gary Illyes:** [00:05:03] For example, you take the robots.txt disallow rule as a way to block something from Search or remove something from Search. And then let's say that that page is very, very popular on the Internet, so we still want to index the URL of that page. And then your expectations are not met because you're still going to see that URL in Search results, not the content, like the URL will be without the snippet, but the title link will still be there.

[00:05:33] And that's because sometimes that can happen. It's very rare, compared to the rest of indexing. It's like the robot URLs in our index is tiny, tiny, tiny minority, but it can still happen nonetheless. And then if you want to remove that URL, then you would think that you can just put up a meta noindex, meta robots noindex in the HTML, in the HTTP header, but then we would never see that if the robots.txt disallow is there.

[00:06:07] **Lizzi Sassman:** [00:06:07] I see. So knowing when these things happen can help you sort of troubleshoot or like you're now aware that like, "Hey, Google may not see this thing that I'm putting up on my

page, because I have this first thing that's sort of blocking earlier on in the process."

[00:06:22] **Gary Illyes:** [00:06:22] Yeah!

[00:06:23] **John Mueller:** [00:06:23] So robots.txt is basically all about crawling? Now that I think about it, like it has a Sitemap file and where you can and can't crawl, so it's basically for that.

[00:06:34] **Gary Illyes:** [00:06:34] Yes. That's also one thing that we work quite hard when we were open sourcing our robots.txt parser and matcher. We removed certain things from the parser, and we also blogged about it, that didn't make sense from crawling perspective.

[00:06:50] So, for example, it used to be that and I know that externally, this is still a pain point, but noindex was there. You could use a noindex column and then path. But we really needed-- like search engines really needed to have robots.txt just for crawling. And then you can control with other things indexing and serving, like generating search results, but robots.txt should just be about crawling and nothing else.

[00:07:17] **John Mueller:** [00:07:17] So cool. Now, I never thought about it that way, but makes total sense. Also the Sitemap file, kind of mention that there, then kind of seems like the right fit. So maybe it should be renamed into crawling.txt.

[00:07:31] **Gary Illyes:** [00:07:31] No, no, no, no, no renaming, no. You know, that for the past, I think, three or four years, we've been working with Lizzi and Henner, one of the engineers I was working on the Googlebot team, on standardizing the robots.txt protocol together with the original author Martin, and it feels like it's kind of a pain.

[00:07:55] **Lizzi Sassman:** [00:07:55] [laughing] Any change suggested. I think the more changes we're like, "Oh no."

[00:08:00] **Gary Illyes:** [00:08:00] Yeah, exactly. Especially now that we are in the finish line of that process, and every single time I see a comment coming in that. "Oh, you need to update this." It's like "No, just no. What? No!"

[00:08:14] **Lizzi Sassman:** [00:08:14] But then if you don't get comments, you're like suspicious, like, "Did you look at the thing?" Because there have been so many comments before. If there is none, it's like something's not right here, potentially.

[00:08:25] **Gary Illyes:** [00:08:25] I think that's just you.

[00:08:26] **Lizzi Sassman:** [00:08:26] That's just me? If I just like the change, let's say, you sent me. If I just LGTM'd, no comments. You would be like, "Yeah, good," or would you be suspicious?

[00:08:37] **John Mueller:** [00:08:37] Gary would just publish it, and it's like, "Okay, cool." [laughter]

[00:08:42] **Gary Illyes:** [00:08:42] I mean, I'm sending it to you for comments. I know that you are going to leave comments. Like, I could have sent it to Adu, one of the engineers who we work with on the parser, and then he would have approved it, but then, I would have gotten no improvement comments.

[00:08:56] **Lizzi Sassman:** [00:08:56] We would have gotten improvement comments by these other people, the committee--What's the term for them? The group that's reviewing the stuff, IETF...

[00:09:07] **Gary Illyes:** [00:09:07] Directorates.

[00:09:08] **Lizzi Sassman:** [00:09:08] Directorates, yes. They pay close attention to everything. And I try to anticipate, and still, we get comments.

[00:09:15] **John Mueller:** [00:09:15] I think it's a good time for me to bring up the suggestion to rename it to crawling.txt then, right?

[00:09:20] **Lizzi Sassman:** [00:09:20] Oh no, in the open-- No, we're closed. We've past that stage, I think.

[00:09:27] **Gary Illyes:** [00:09:27] Yeah. So at this point, we only get the comments from the directorate, live from the Security Directorate or the area review team or whatever, and there are no more community contributions. Which is great because fewer comments, I guess. But yeah, we've been dealing with

comments for a long time and I would rather not introduce a change into the protocol.

[00:09:50] **Lizzi Sassman:** [00:09:50] Would that mean that people would have to change all of their files, like, old named robots.txt files? Everything would break, like it would be chaos.

[00:09:57] **John Mueller:** [00:09:57] Okay, no chaos. No chaos. We can...

[00:10:00] **Gary Illyes:** [00:10:00] We like it. We thrive in chaos. We love chaos.

[00:10:04] **John Mueller:** No.[00:10:04]

[00:10:05] **Gary Illyes:** [00:10:05] No, but so, technically, it wouldn't change anything, because it's down to the implementers to support the things that we mentioned in the specification. So, for example, if they still want to follow the old naming like robots.txt, then they could just do that. Like, it's not like it's set in stone. And if you look at HTTP standard, for example, you can find things that browsers in general don't support, even though it's in the specification.

[00:10:30] So it's not like, you must absolutely do this. It's like here are the considerations that we thought of. We think that you must follow this and this. You should follow this and this. But ultimately, it's up to you what you do with it. We already talked a lot about robots.txt, so I would rather not make this an episode about robots.txt again.

[00:10:50] **Lizzi Sassman:** [00:10:50] OK, so let's return to the crawling bucket. What other things do we notice in the crawling stage that it's related to a removal? Like what other things could we notice at that point right away that "Hey, somebody has requested this thing be removed." So there's robots.txt, what other things would be noticed at the crawling point?

[00:11:14] **John Mueller:** [00:11:14] I guess the URL removals tool used to be a thing there. Where if you flagged the URL parameter as something that we should skip, then that would also be dropped. But that tool doesn't exist anymore. So probably nothing. Or can you think of anything, Gary?

[00:11:32] **Gary Illyes:** [00:11:32] I'm thinking about status codes.

[00:11:34] **Lizzi Sassman:** [00:11:34] Oh, because that's the other thing, if the robots.txt file is like a 404 or something or like we can't access it, then we don't know what to do. So that page itself could be...?

[00:11:47] **Gary Illyes:** [00:11:47] Well, not 404, because that means it's missing. Like 503 or 505 or whatever, but not even just that. Like, for example, if you return a URL with a 404 status code, then that would be removed from Search. It's not a removal tool. Like, we are not saying that this is a tool for removing URLs, I don't think we do. But technically it works, and it's just working as intended. Basically, you instruct the client with something. Well, you don't instruct, you inform the client.

[00:12:15] **Lizzi Sassman:** [00:12:15] Well, we do say one way to remove URL would be to delete it. So then you would serve the status code. So then that would be working as intended as like, "Hey, it's gone, cool! Good job. It's out." So we would have to crawl it and then decide not to index it.

[00:12:30] **Gary Illyes:** [00:12:30] I mean, if it's 404 or 410 or 403, maybe then, we just wouldn't even pass it on to indexing. We would pass on a signal to indexing that like, "This URL doesn't exist any more. " Then indexing would decide that we are going to drop it.

[00:12:44] **Lizzi Sassman:** [00:12:44] OK.

[00:12:45] **John Mueller:** [00:12:45] So, what else happens in the indexing bucket?

[00:12:48] **Gary Illyes:** [00:12:48] Not much!

[00:12:49] **Lizzi Sassman:** [00:12:49] (laughing) This is when we would find like all those other tags like noindex meta tag, no-follow, like things that are on the page. Don't go here. Don't go over here. Stop looking at me.

[00:13:01] **Gary Illyes:** [00:13:01] I think in an earlier episode of Search Off the Record, we talked about indexing a little bit, and one of the things that we mentioned was the content parser and rendering. And during those processes we do extract the meta tags that are relevant to indexing.

[00:13:17] So for example, rel="canonical" is extracted, meta robots is extracted and so on. And if we see something like meta robots noindex. And if we see meta robots noindex, then that will mean that we want to drop that URL from the index.

[00:13:33] **Lizzi Sassman:** [00:13:33] What would happen if you've got other languages? Like you've got the noindex on one language, would we then do something about the other languages? Like, if, I guess, you set it as the rel="canonical" , then?

[00:13:48] **Gary Illyes:** [00:13:48] So it depends on how the setup is, because it can happen that you, for example, if we have some sort of duplication across languages and Google detects that as a duplication than it might mean that one noindex can affect the whole cluster. But it's hard to trigger it, but nonetheless you can.

[00:14:07] **Lizzi Sassman:** [00:14:07] OK, so, it's something that could happen. But I'm wondering, is it safe to rely on that? Or you should put it on every language page, just to be sure if you wanted it for sure noindex is removed? You should be thinking about all the things. Or can you just add it to the canonical page and then we will know to get rid of it? Like from just an efficiency perspective, I guess. Like, I don't want this page, so I therefore also wouldn't want the other pages that I've linked up.

[00:14:34] **Gary Illyes:** [00:14:34] You probably want to add it to every page that you want to remove. I think our setup is more special, because we do use hreflang. So, we are instructing--not instructing, but hinting to indexing that these are clusters of pages, and then they might influence each other.

[00:14:54] **Lizzi Sassman:** [00:14:54] Might. I mean, I know it sounds like maybe, and if you wanted to be 100% sure, then you would...

[00:15:02] **Gary Illyes:** [00:15:02] So that's why I'm saying that we are more special, like pages with hreflang clusters are more special and you have to keep in mind that they might affect each other in the cluster.

[00:15:12] So for pages that are not in a cluster, not in a hreflang cluster, for example, there you can just noindex the pages that you want to get removed and don't think about anything else. Just remove the pages or noindex the pages that you want to remove. With hreflang clusters it gets more complicated. Also, recurring topic that hreflang is complicated. Yes, it is.

[00:15:36] **John Mueller:** [00:15:36] So how does rendering fit into this? If you add the noindex robots meta tag with rendering in JavaScript? Would that still work? Or is that a bad thing?

[00:15:45] **Gary Illyes:** [00:15:45] We are trying to figure out what happens with rendering, but I'm thinking, in theory, at least in most of the cases, it should work. But we definitely need confirmation from Martin about what's actually happening.

[00:15:59] **Lizzi Sassman:** [00:15:59] And the reason you bring this up is because it's, "Are we going to see it in time?" If it's injected with JavaScript, would we know to not show it, because we haven't seen it?

[00:16:10] **John Mueller:** [00:16:10] Of course, kind of like the timing difference. There might be a window in between. Or if we don't render a page for whatever reason once, then what happens?

[00:16:18] **Lizzi Sassman:** [00:16:18] Right, like we didn't know that, "Hey, this is actually noindex, because we didn't see it when Googlebot is rendering it."

[00:16:25] **John Mueller:** [00:16:25] OK, so, basically, be as consistent as possible, just like if you wanted noindex, just noindex takes everywhere.

[00:16:33] **Gary Illyes:** [00:16:33] Yes, I'm trying to think of like corner cases, but I can't come up with anything useful.

[00:16:39] **John Mueller:** [00:16:39] How could you put the meta tag on things like PDFs?

[00:16:43] **Gary Illyes:** [00:16:43] Well, you can't.

[00:16:44] **John Mueller:** [00:16:44] Oh, that sucks.

[00:16:46] **Lizzi Sassman:** [00:16:46] [laughter]

[00:16:47] **John Mueller:** [00:16:47] What can you do there? Like you have a PDF, and you're like, "I just don't want this indexed"?

[00:16:51] **Lizzi Sassman:** [00:16:51] Or you want to removed, basically. Like, it's being served in Search results and you want it removed, how would you go about-- I guess you just delete it?

[00:17:01] **Gary Illyes:** [00:17:01] We do support x-robots-tag HTTP header. And then for binary files, you can just add that, with a noindex thingy. And then that will work like a meta noindex.

[00:17:13] **Lizzi Sassman:** [00:17:13] OK, so this method, the header method. Which bucket does this fall into? Would we notice it at the crawling or at the indexing stage?

[00:17:23] **Gary Illyes:** [00:17:23] Technically, if we would notice it in crawling and then we forward it to indexing, and then indexing will drop it. So, that's the signal that we sent with, like, the 404. So, basically, we create a wireframe of whatever we collected about the URL, the timings and different meta information that we could collect including the headers. And then based on that, indexing can decide what to do with the page.

[00:17:49] So basically, in crawling, we just notice it, but we don't actually do anything. It will be indexing that will have to do something about it, and same goes for status codes as well.

[00:17:59] **Lizzi Sassman:** [00:17:59] OK. And when you say do something about it, what does that mean?

[00:18:02] **Gary Illyes:** [00:18:02] Keep it or drop it.

[00:18:04] **John Mueller:** [00:18:04] So it's basically a blockchain from...

[00:18:06] **Gary Illyes:** [00:18:06] No, no, don't call it that. Bad John.

[00:18:11] **Lizzi Sassman:** [00:18:11] You can just join up any words together. [laughter]

[00:18:14] **John Mueller:** [00:18:14] I'm German. In German, you can just take random words and put them together and make a new word.

[00:18:19] **Lizzi Sassman:** [00:18:19] What if they're existing words? This is a term already.

[00:18:23] **Gary Illyes:** [00:18:23] Then you created a pun.

[00:18:06] **John Mueller:** [00:18:06] OK, so moving on to the third bucket. We can call it Web3.

[00:18:30] **Gary Illyes:** [00:18:30] No.

[00:18:31] **Lizzi Sassman:** [00:18:31] OK. [laughter] It's like he's not hearing us, on purpose! Selective listening.

[00:18:40] **Gary Illyes:** [00:18:40] It's a pleasure working with you, John. [laughter]

[00:18:42] **Lizzi Sassman:** [00:18:42] I have actually a question regarding the crawling bucket. So in the docs, we have unavailable after-- I don't know if we talked about this last episode. But there is a part in there that says, if you add that, then crawling will slow down over time, which is interesting I guess.

[00:19:04] **Gary Illyes:** [00:19:04] Is it, though?

[00:19:04] **Lizzi Sassman:** [00:19:04] Yeah, I mean, so you put it there like you want it removed, but we're just slowing down the crawling, we're not stopping completely? Or like we still check back on it, we don't see it as a "Hey, remove it." It's sort of a soft exit like an opposite of a soft launch.

[00:19:22] **Gary Illyes:** [00:19:22] So I don't actually know how exactly it works or how it's implemented, but my understanding was always that it's a soft exit, because we still need to ensure that your not bull... [bip].

[00:19:36] **Lizzi Sassman:** [00:19:36] Wait, but I put it there like whoever it is, who put that tag, like, I guess...

[00:19:41] **Gary Illyes:** [00:19:41] No, not the tag,

[00:19:42] **Lizzi Sassman:** [00:19:42] You don't believe then that the content is actually available after...

[00:19:45] **Gary Illyes:** [00:19:45] Like with the last mod tagging in Sitemaps.

[00:19:50] **Lizzi Sassman:** [00:19:50] Oh, I see.

[00:19:45] **Gary Illyes:** [00:19:45] That's also something that we might use it. If we can detect that it's consistently accurate.

[00:19:58] **John Mueller:** [00:19:58] OK.

[00:19:59] **Gary Illyes:** [00:19:59] My understanding is the same goes for unavailable after.

[00:20:02] **Lizzi Sassman:** [00:20:02] And is this something that people would use for sort of schedule a removal or something? Like they eventually want it to be removed from Search results, so they put it there to sort of preemptively, like, "Hey, this thing won't be relevant. but it could still hang out in Search, just crawl me less and if it falls out eventually, that's fine with me."

[00:20:21] **Gary Illyes:** [00:20:21] I think so. John is aggressively nodding here.

[00:20:24] **Lizzi Sassman:** [00:20:24] Yeah, but he's also sort of making a face. I'm not sure. Do you concur or...?

[00:20:29] **John Mueller:** [00:20:29] Yeah, I mean my understanding was, we understand from the page that they want to have it go away at some point. And then, either we can just blindly trust it and start hiding that page in Search, or we double check it at some point and then it's more like a noindex and then whatever happens there.

[00:20:48] **Gary Illyes:** [00:20:48] From maybe a decade ago, some auction site like big auction site was using this, and they were using it accurately and consistently. And you could actually see the auction pages disappearing from search when the auction expired. It might have taken a day or two, because we actually needed to crawl the page again, but it actually worked, if nothing else for scheduling the crawl to that date.

[00:21:15] **John Mueller:** [00:21:15] So, going back to the last bucket. I won't mention its name. What kind of blocking could you do there? So basically, it's basically indexed in Google systems.

[00:21:27] **Gary Illyes:** [00:21:27] I mean you can do what Lizzi said. Basically removing altogether. But you can also control the snippets that we show.

[00:21:34] **Lizzi Sassman:** [00:21:34] Would this be also like removing the cache? Like the cache gets removed at the serving stage or no? Like we still have it but then...

[00:21:43] **Gary Illyes:** [00:21:43] That's kind of weird. I think it's good that we have the cached page, because when the pandemic hit, then we could serve some pages from there when the site went down due to heavy load. But I usually don't even think about the cached page. But yeah, you can remove from the cached page. For example, whenever we have there, you are not removing from search results anything, you are just removing the cached page. So you are controlling the cached page, not the content that we might have indexed.

[00:22:15] **John Mueller:** [00:22:15] So that's the noarchive meta tag, right? There's also the cache removal tool.

[00:22:20] **Gary Illyes:** [00:22:20] Right.

[00:22:21] **John Mueller:** [00:22:21] Where we used to have one.

[00:22:22] **Lizzi Sassman:** [00:22:22] And that would be for what purpose? Like, you would use... What do you mean we used to have it. We don't have it anymore?

[00:22:28] **John Mueller:** [00:22:28] I don't know. I haven't taken a look for a while, so I was kind of like, "Do we still have that?" Maybe one of you knows.

[00:22:34] **Lizzi Sassman:** [00:22:34] Yeah, I'm wondering, can you use the removals tool to remove the cached page, or like, what's in the cache? Or would that stomp on the full page too? Like, "Whoops, we removed the whole thing." Like, you just wanted us to remove the cache and there's an updated version, but if you put the cached page in there, would we...

[00:22:51] **John Mueller:** [00:22:51] Let me assume this tool still exists, because probably it does. Otherwise, someone would have told us, hopefully.

[00:22:59] **Gary Illyes:** [00:22:59] Or we will get complaints about it.

[00:23:01] **John Mueller:** [00:23:01] Or we would get complaints about it. So my understanding is there are two ways that you can remove a page with these tools. One is your remove it completely if you're like, "I don't want this page at all to be visible in Search." And the other is it removes the snippet that is shown and the cached page, which is like, "I changed something on this page and I want Google to..."

[00:23:26] **Lizzi Sassman:** [00:23:26] Refresh it, like show...

[00:23:27] **John Mueller:** [00:23:27] Refresh it and definitely don't show the old one for the time being, kind of thing. Like you put a wrong phone number on a page and people are calling your neighbor instead of you, you can say like, "Fix this as quickly as possible or at least hide it from people seeing it."

[00:23:43] **Lizzi Sassman:** [00:23:43] I guess, what most people think, like recrawl the URL, not remove the cache, it's sort of a like a way to think about the same thing but maybe not typical.

[00:23:52] **John Mueller:** [00:23:52] I think removing the snippet in the cache page is more for people who are not the site owner. So, for them, they can't go to Search Console and say, "Hey, recrawl this random person's web page for me please." But they can go and say, "Actually, the information on this page has changed, I need you to reflect that in Search as quickly as possible."

[00:24:14] **Gary Illyes:** [00:24:14] It's quite an important aspect. I think that if someone wrote something about you and it was not accurate, then you definitely want it to be gone from Search once they update the article or whatever.

[00:24:30] **Lizzi Sassman:** [00:24:30] And is this happening in the rendering bucket?

[00:24:33] **Gary Illyes:** [00:24:33] No, we are already in serving bucket.

[00:24:35] **Lizzi Sassman:** [00:24:35] We are already in serving, oh!

[00:24:36] **Gary Illyes:** [00:24:36] We left you behind.

[00:24:38] **Lizzi Sassman:** [00:24:38] Oh, whoops. [laughing]

[00:24:39] **Gary Illyes:** [00:24:39] And then we have the normal removal tool which you can actually use to remove a whole result from search results.

[00:24:48] **Lizzi Sassman:** [00:24:48] And from our perspective, is that like going through all the buckets again? Or does it like jump start you?

[00:24:54] **Gary Illyes:** [00:24:54] So that's removing from Search result, like that the phrasing was important there, because it is just hiding from Search results in the first step. And if you're a site owner and you want something gone from Search completely, like from the index and from Search results, then you also have to take one of the previous actions.

[00:25:15] Like, for example, I add a noindex or--just add a noindex as the best one. Or remove completely, like 404 return, 404 or 410. But otherwise, the removal tool, that will just remove the search result itself.

[00:25:29] **Lizzi Sassman:** [00:25:29] Oh, is that why it's so fast or it's the faster method?

[00:25:33] **Gary Illyes:** [00:25:33] Yes, that's correct. And that's also why we have it, because in some cases, at least we cannot refresh the index that fast.

[00:25:40] **Lizzi Sassman:** [00:25:40] Right!

[00:25:41] **Gary Illyes:** [00:25:41] It might sound really weird, but sometimes it may take months to remove a result from our actual index.

[00:25:49] **John Mueller:** [00:25:49] Then, does it just get removed from the data center that the serving is happening at? Or does it drop out when you search?

[00:25:57] **Gary Illyes:** [00:25:57] So it drops out when you search. So basically if the removal tool was used on a result and the result is still in the index, the URL is still in the index, then we would normally do the retrieval that we do for any other result. But then, we would not present that result to the user, to any user.

[00:26:21] **Lizzi Sassman:** [00:26:21] But it doesn't necessarily flag the page to not be crawled again. So we could go visit it and see...

[00:26:26] **Gary Illyes:** [00:26:26] Well, you want to crawl it again.

[00:26:28] **Lizzi Sassman:** [00:26:28] Right, but say the person uses the removals tool, that doesn't prevent it from us finding it again. They would have to do some other thing, because we're still going to attempt it.

[00:26:38] **Gary Illyes:** [00:26:38] That's correct. That's why you want to use noindex or 404 or whatever. And then that would permanently remove-- not permanently, because if it comes back or the noindex is dropped, then it would come back. But it would remove page that has a noindex or returns with a 404 from our Search index.

[00:26:55] **John Mueller:** [00:26:55] That sounds like a lot of work every time to do, like when someone searches. Because if you search for something and it comes back, and like, "Oh this was done in 0.005 seconds." I guess it has to look through the whole removals file every time. It's like pretty crazy. Wow!

[00:27:13] **Gary Illyes:** [00:27:13] Well, but you loaded in memory. And then if you loaded in memory, like you don't have seek times, you don't have anything. So it's not that bad.

[00:27:21] **John Mueller:** [00:27:21] Is that why regular expressions can't be done there?

[00:27:24] **Gary Illyes:** [00:27:24] John, stop it!

[00:27:26] **John Mueller:** [00:27:26] Regex!

[00:27:28] **Lizzi Sassman:** [00:27:28] That's how I say it.

[00:27:30] **Gary Illyes:** [00:27:30] I'm surrounded by regexers.

[00:27:33] **John Mueller:** [00:27:33] So basically, Google systems are just like face-palming whenever I try to submit something like that.

[00:27:38] **Gary Illyes:** [00:27:38] Well, I'm not sure if there are face-palming. I think they are just assuming that that's part of the URL.

[00:27:43] **Lizzi Sassman:** [00:27:43] [laughing] Very interesting URL structure you have going on.

[00:27:47] **Gary Illyes:** [00:27:47] Yeah, very weird URL. I think theoretically, we could support some form of Regex to some extent.

[00:27:54] But I'm not sure that would do any good. Like, people would wipe out their sites in very or new kinds of imaginative ways.

[00:28:04] **John Mueller:** [00:28:04] I've heard regular expressions are hard. I don't know what they mean with that, but...

[00:28:08] **Gary Illyes:** [00:28:08] Oh so next time I have a Regex problem, I will just go to you, because you're a Regex expert.

[00:28:14] **John Mueller:** [00:28:14] Regex expert, yes.

[00:28:15] ♪ [music] ♪

[00:28:20] **John Mueller:** [00:28:20] Next time on Search Off the Record, we'll be taking a look at accessibility, which is kind of the opposite of blocking, specifically for your site. We've been having fun with these podcast episodes. I hope you the listener have been finding them both entertaining and insightful as well.

[00:28:35] Feel free to drop me a note on Twitter or chat with us at one of the next events that we go to, if you have any thoughts. And of course, don't forget to like and subscribe. Thank you and goodbye.

[00:28:49] **Lizzi Sassman:** [00:28:49] Bye-bye!

[00:28:50] **Gary Illyes:** [00:28:50] Farewell.

[00:28:51] ♪ [music] ♪