

## GDEVS SOTR099 - VIDEO

[00:00:09:23] - **Martin**

Hello, and welcome to a new episode of Search Off the Record, a podcast coming to you from the Google Search team, where we will talk all about Search and maybe have some fun along the way. My name is Martin, and I am a Developer Advocate or Search Advocate at the Search Relations team here at Google. And with me is John. Hi, John.

[00:00:30:01] - **John**

Hi, Martin.

[00:00:31:04] - **Martin**

John, I have a question.

[00:00:32:22] - **John**

Uh-oh.

[00:00:33:25] - **Martin**

I read something, and I thought about it, and I'm not sure what to think about it. Someone said online somewhere that they don't need to do SEO, or they don't need to worry about SEO, because what they do, like the stuff they have on their website, is behind the login. I'm not sure if that means that really they don't have to do SEO, because I think they might still have to do some amount of SEO. What do you think?

[00:01:03:00] - **John**

I think the real answer is it depends.

[00:01:08:10] - **Martin**

All right. Okay.

[00:01:09:13] - **John**

I'm so sorry.

[00:01:11:15] - **Martin**

I'm going to step in here as Barry Schwartz. What does it depend on?

[00:01:15:00] - **John**

I think, if you really don't care about what is indexed, then do whatever you want kind of thing. Like, maybe something will be indexed, maybe it won't, but I have zero care about what is visible in Search. Nobody can access my content anyway, so probably doesn't matter. If you care a little bit about what is visible in Search, then maybe you should think about how you set things up.

[00:01:46:27] - **Martin**

Okay. How would I know how to set it up if I want? I mean, I guess I want my website to show up in Search somehow, but I guess I don't want to just show like the login page, right?

[00:02:01:03] - **John**

Yeah, I mean, there are a variety of different directions that you could go there. The variations I usually see are things like paywalled content, where basically you do want Google to index things, but the content itself might be behind a paywall or a login page or something like that. So, when a user comes, they would see the interstitial to log in. We have a bit of documentation on how to set up paywalled content. Perhaps that's not really what the person that was asking you was asking about, because it also sounds like they don't want to show the content to Google either, which is fine.

[00:02:51:29] - **Martin**

Hmm. Yeah.

[00:02:53:27] - **John**

With paywalled content, what usually happens is you try to recognize when Google is crawling and you serve Google the content that you want to make available, and you add the paywall structured data to the page to make it clear to Google that, "Hey, actually, this content is not available to everyone. There's some limitations," and that could be maybe you require a login, maybe you require a payment, maybe after a certain number of iterations you're like, "Oh,

this is enough free content." Now you have to pay for it. There are lots of variations with regards to paywalled content. It also doesn't have to be something that's behind a clear payment thing. It can just be something like a login or some other mechanism that basically limits the visibility of the content.

[00:03:49:12] - **Martin**

Ah. For instance, if I have to like watch a video or click on an ad or something to get to the rest of the content, that's kind of also fine.

[00:03:57:10] - **John**

I don't know about fine, but that kind of falls into the category of, "Well, there's something that needs to be done before this content is actually visible." And then you would use a paywall structured data.

[00:04:10:10] - **Martin**

Okay.

[00:04:11:14] - **John**

Also, if you have something like different thresholds where you say some people get to view five pages for free and others have the whole content available for free because you're doing A/B testing maybe about the prices or things like that, then you'd want to use a paywall structured data, just to make sure that when Google is looking at it, they realize that sometimes this content is not available.

[00:04:37:18] - **Martin**

Okay. Got it, got it.

[00:04:40:01] - **John**

The paywall structured data helps us to understand that users might see something different, and that's totally fine. I think there's one thing maybe to watch out for with paywalled content is that, when a user looks at your page, you don't load the content into the HTML, but rather you make sure that it's really not loaded into the page's DOM so that, if a browser has something like, what is it, the text reader or speech reader.

[00:05:14:05] - **Martin**

A screen reader? Yeah.

[00:05:15:15] - **John**

Screen reader that the screen reader doesn't go off and read all of this text that you're trying to hide. Those kind of things. That would be kind of the thing that I would watch out for that, if it's really paywalled content or limited content, make sure you don't load it into the browser and use JavaScript to turn it on, but rather that it's really only served to the user when you want to make it available.

[00:05:40:02] - **Martin**

Okay, but that's like one specific kind of content that you are hiding away or making not immediately accessible. But what if I have, I don't know, a website where I share apartment ads, for instance, or apartments to rent, and I want people to log in to see the apartment, to interact with the apartment, or to apply for the apartment. How would I go about that? Is that just like immediately show a login page? Or are there better ways of doing that?

[00:06:16:07] - **John**

I guess the question would then be, do you want this content to be visible in Google or not? If it's visible in Google, then that would be kind of the model of paywalled content. If you don't want it visible in Google, maybe like this is a private forum or a private community where you're sharing things, or maybe you have something like, I don't know, a private service where people who have a subscription, they have access to this content, but it's not shown in Google or in specific tools or something like, I don't know, you have a spreadsheet that runs in a browser kind of thing where everyone has their own private content and they all have URLs.

[00:07:00:07] - **Martin**

Okay, fine. All right. How are bigger services doing it? Do you know if they just show the login page or how do they do this? I don't think they use all paywall structured data.

[00:07:19:01] - **John**

Yeah. I think, if you're looking at a service like, I don't know, Search Console or Google Drive where you have kind of this private content that is hosted online with a specific URL, then fundamentally, in order for someone to see that content, they have to log in. Usually I guess, it depends on how they set it up. But, oftentimes, when you try to access a page like that, it'll redirect you to a login page, and I think how they deal with the login page determines a little bit

how things could potentially end up being indexed. For example, with Search Console, one of the things that they do is that they have a set of marketing pages that are freely available. If you try to access a Search Console URL directly without being logged in, it'll redirect you to a marketing page, which has a link that says, "You can sign in here," to actually get the full information.

[00:08:29:02] - **Martin**  
That makes sense.

[00:08:30:04] - **John**  
I think, from an SEO point of view, that's fantastic because you search for Search Console and you can find these marketing pages. If anyone accidentally links to their private Search Console URL--like "Oh, I want to look at the performance report for my site," they share that with people--then that URL will redirect to the marketing page so that, on the one hand, users who find that link randomly, they end up on the marketing page, they know what it's about. And, for search engines, they will find this marketing page and they'll be like, "Oh, okay, this has indexable content. We will just index this."

[00:09:11:22] - **Martin**  
Okay. We are seeing it depends a little bit on what kind of content you're hiding and there are in between bits and pieces. Like, you don't have to just completely direct to a login page, I guess. Okay. Interesting.

[00:09:26:11] - **John**  
One of the things we noticed over the years, specifically around login pages, is that if you have a very generic login page, we will see all of these URLs that show that login page, that redirect to that login page, as being duplicates. Like, if whenever you access a private URL, it just says username and password, then we will think all of these individual private URLs are actually the same. We'll fold them together as duplicates, and we'll focus on indexing the login page, because that's kind of what you give us to index. That means in the search results that login page is going to be very popular because all of these random links, they keep redirecting to it or they keep showing the same login page. If someone is searching for your service and they want to know more about your service, and the only thing or the primary thing they find in search is like, "Here's how to log in," that might be a kind of a weird experience for them.

[00:10:31:11] - **Martin**  
Okay. That is, yeah. I mean, yeah, that's not great. Should they then, for instance, check if it's a legitimate Googlebot and then just give the actual content of the URL, or at least like some sample content? Or how would you fix that specific problem where everything gets deduped?

[00:10:51:17] - **John**  
I think, if this is private content, you don't want to share that with Googlebot.

[00:10:57:28] - **Martin**  
Well, okay. If it's private content, sure. If it's private content, you don't. But then how do you keep Googlebot from putting it in the index. You just put a noindex on it? Or robot the URL away so that we're not even crawling it? What's the idea?

[00:11:13:11] - **John**  
Okay. I think for these situations where you want to show a login page, it's good to have some context on the login page. The Search Console model is basically show a marketing page instead of the login page. But, if you have a generic login page, put some information about what your service is on that login page, which could be enough to just have a sample of text like, "Oh, you're accessing Martin's Furniture Lookup site," or, I don't know, some intranet thing where maybe some private content is. And then, if you have some information on that login page, then we can index that that information. If you have different types of services that use the same login page, then those different services will have slightly adjusted login pages. If you're searching for, I don't know, maybe we'll just stick with Google Drive. If you're searching for Google Docs, you'll find a login page, maybe, for Google Docs. If you're searching for Google Sheets, you'll find a login page, maybe, for Google Sheets. Having a little bit of information on there is important. The other thing you mentioned is whether all of this should just be blocked by robots.txt, which is another common strategy for dealing with things that you don't want to have indexed. The problem, I think, with doing that is the URLs could become indexable so we wouldn't see the contents of the login page, but rather we will just see like, "Oh, people are linking to this specific Google Doc and we can't access it, but maybe we should show it in the search results if someone is searching for something similar." Also, this could be visible if someone does something like a site: query for your site and it's like, "Oh, tell me all of the URLs that are indexed for this hidden section of a website," and then Google and other search engines might be like, "Oh, I know about all of these URLs. I don't have any information on what's on there, but feel free to try them out," essentially, which is probably a bad idea. If you have random hashes in the URL, so a collection of random characters, it's not a bad thing or not a terrible thing. But, if you

have things like usernames or email addresses in the URL, then of course all of those could become indexable. So, if it's private content, serve it with a noindex or redirect it to a login page somewhere. Don't use robots.txt.

[00:14:00:05] - **Martin**

And, ultimately, don't leak private details in URLs.

[00:14:03:19] - **John**

Sure. Yeah. Of course. Yeah, I think that's always a good practice. But sometimes you have things like you form submission parameters in a URL somewhere, and it gets stuck there.

[00:14:19:29] - **Martin**

Okay. Any other common problems you're seeing with login pages specifically, or with content that is behind some sort of login?

[00:14:30:11] - **John**

Yeah, I think the other question that I sometimes run across is whether or not the login page should be indexable by itself, and I think that depends a bit on the nature of the content that you have behind the login page. For example, if you have a kind of an intranet that is available publicly where your employees can only access it, then you probably don't need that login page indexed in search because like your employee should be able to find the URLs for your private content on their own. Hopefully. In a case like that, you probably can just serve, I don't know, the login page with an error code, or use server-side authentication, or put a noindex on the login page so that, if it does get found, then at least it won't be indexable like that. I think one aspect as well, which I've seen in the past, every now and then, that people's intranets end up getting indexed. There's a login form, but you probably don't want people to accidentally run across your intranet URLs. Yeah, I think those are kind of the primary aspects, and showing a login page is generally fine. Whether or not you redirect to a login page or show the login page directly, ultimately I think is more a technical decision on your side. Sometimes there are security implications around that.

[00:16:08:21] - **Martin**

Whoa, security implications.

[00:16:11:00] - **John**

Well, I think cookies, for example, right?

[00:16:14:01] - **Martin**

Ah. Okay. Fair.

[00:16:15:08] - **John**

So maybe you have something like login.yourdomain.com and everything gets routed through there, then you want to redirect to login page there.

[00:16:24:23] - **Martin**

That makes sense. Yeah. Okay. Yeah, sure. Okay. I see what you mean with security implications. Okay.

[00:16:31:04] - **John**

I think this is a problem pretty much for any site that has kind of private sections on the site which are accessible through individual URLs, but definitely a problem for sites like Google Drive or all of the various Google services where you end up having a lot of content that is private to yourself, to the user, and where you have a lot of different login pages. Specifically, if you have multiple services that go through the same login page, then it's worthwhile to think about how you actually want your service to be findable in the search results. For the most part, you do want things findable. If people link to something private, you do want something smart to happen there. So it's kind of good to think about how you should combine things, and we regularly see Google services getting this wrong--or getting, I mean, not necessarily getting it wrong and that you can access the private content, but wrong in the sense that, we index things that probably we shouldn't be indexing like that.

[00:17:54:05] - **Martin**

And then all you get is a login page. Yeah. That's not great.

[00:17:57:11] - **John**

Yeah, yeah. I think Search Console used to have that problem before they moved to kind of having the marketing pages as a redirect target where you would search for Search Console and you would find someone's Search Console URL in the search results, and it's indexed as "Sign in here" kind of thing, which is like it's a login page. Of course you can reach Search Console that way, but it's not really the best way to show Search Console in the search

results. And, because Google has so many different services and so many different teams working on these services, you invariably run across situations like that.

[00:18:40:00] - **Martin**

I mean, for some of the services is also tricky. If I have a Google Doc that I make public, kind of like a non-website website, so to speak. And then it gets indexed, and then it is visible and people actually can use the content. Then I delete the file or if I make it private again, then it is indexed. It will take a while until it falls out of the index. So there will be surprises. Let's put it that way.

[00:19:04:19] - **John**

I mean, surprises in the sense that if you're not prepared. Sure. But I think it just makes it hard for search engines to go and actually index or find content on Google Docs where it's like, "Oh, maybe there's something here, maybe all of this is private, probably it's private, but maybe I should check anyway."

[00:19:27:06] - **Martin**

Yeah.

[00:19:28:28] - **John**

Yeah. I think the other aspect that's kind of interesting is that internally we don't give SEO advice on these kind of things. Every now and then, someone with a public service will ping us internally at Google and be like, "Oh, how do I make sure my service is indexed properly?" Essentially, we have to point them at our public documentation. Maybe we'll point them at this podcast in the future. Yeah. But it's something that just comes up every now and then. I think larger websites, especially those that have private content, they probably have similar things. Even e-commerce sites, where you have something like you can look up your account or the orders that you had in the past, they will have a specific URL and maybe someone will link to that and a search engine will try to index it. How you handle that kind of depends on what is actually shown in the search results.

[00:20:29:12] - **Martin**

Yeah. And whatever makes sense for a user who might land on that or want to land on that. Yeah, that makes sense. All right. Okay. Would you say there's something that people should do to make sure they are doing this right for them? Is there like the top tip that you want everyone to take away from who has to deal with logins?

[00:20:54:02] - **John**

I think the most important part is that you understand how things are currently working for your site. The way I would do that is I'd open an incognito window in a browser where you're basically not logged in to any of the services that you usually use, and then you search for something associated with your site that could be, like if the primary content is behind a login page, then you could search for your name, like the name of the service. You could search for, I don't know, Search Console or Google Docs or something like that. And then you click on maybe the top couple of results to see what actually comes up there. If the top result is something like a login page and there's no information on this page at all otherwise, then probably that's something that you can improve. Whereas, if the top results are kind of reasonable marketing content for people who are not logged in yet, then that seems okay. I think, with regards to more specific sections of a site, that gets a little bit harder because you have to search for those parts specifically, you almost have to know that there's something that could be found. For example, on an e-commerce site, if you have a page that shows your orders, you could search for that URL pattern or specific words that might be on a page like that and see what comes up. From there, while you're not logged in in an incognito window, see, is there actually reasonable content that comes up? Does it do a reasonable kind of a redirect to a login page or, I don't know, login experience if you want to add more content to those pages? Or is this kind of jarring for the user? That is like, what am I doing here? Why did I end up on this page that's asking for a password now? I'm not trying to hack this website kind of thing. That's kind of the direction I would go there. If you see that things are okay in the search results, then probably you're already doing things properly. If you see that things are not going okay, then I would recommend digging into those specific URLs, trying to figure out where are they coming from. What happens when you use Search Console's URL Testing Tool to look at those pages. Does it show what you see or does it show something different? Based on that, you would try to make a plan for improving things.

[00:23:35:27] - **Martin**

Okay, that sounds pretty good. I think that's pretty actionable advice, especially with checking how your service currently presents in Search to someone who's not logged in is probably a very, very good first stage to make sure that you have a good customer experience in the end. So that makes sense. All right. I think pretty much that's sorted, huh? Do you have anything else you want to say about login pages?

[00:24:04:20] - **John**

I could tell you more, but you have to log in first, Martin.

[00:24:10:03] - **Martin**

How does that show up in the index? Give me some sample content first. I want to know if I want to. Is that a paywall? Do you want me to pay for this information?

[00:24:21:12] - **John**

No, but you should subscribe to this podcast, and then we'll tell you more.

[00:24:26:20] - **Martin**

And that's actually free, so definitely do subscribe. Leave us a comment. Have you seen any services that are screwing this up in the search results, or do you have more questions regarding login pages or paywalls? Let us know in the comments below. We'll probably be talking about these specific issues more in depth. You can also submit to the Office Hours as well if you have a specific question. But, if it's a broader thing, then we might discuss it here in the podcast. Awesome. Well, John, it has been a pleasure. Thank you so, so much for being here. I think I've never thought this much about login pages, I don't know. For me, they are always just like username and password or email and password and then like a button and that's it. But, yeah, there's more to it. Thanks a lot for joining me. Thanks to all the listeners out there. That's it for this episode. I do hope people enjoyed that a lot. John, if they want to talk to you, where do you hang out online these days? Behind or not behind the login? Where can people reach out to you?

[00:25:28:01] - **John**

I don't know, sometimes it's hard. I'm mostly active on Bluesky nowadays, so people can drop me a note there or send me a private message if they log in. That would potentially be a good place.

[00:25:43:14] - **Martin**

Okay, so everyone follow John on Bluesky, and thanks a lot for listening. Please do like and subscribe if you enjoyed this episode, and goodbye.

[00:25:53:15] - **John**

Bye.

[00:25:57:21] - **Martin**

We've been having fun with these podcast episodes, and we hope that 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 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.