SMX replay: SEO that Google tries to correct for you

Search engines have seen the same SEO mistakes countless times, and as Patrick Stox, SEO specialist at IBM, said during his Insights session at SMX Advanced, “Are you going to throw millions of dollars at a PR campaign to try to get us [SEOs] to convince developers to fix all this stuff? Or are you just going to fix it on your end? And the answer is they fix a ton of stuff on their end.”

During his session, Stox outlined a number of common SEO responsibilities that Google is already correcting for us. You can listen to his entire discussion above, with the full transcript available below.

For more Insights from SMX Advanced, listen to Amanda Milligan’s session on leveraging data storytelling to earn top-tier media coverage or Ashley Mo’s session on improving your YouTube ad performance.

Can’t listen right now? Read the full transcript below

Introduction by George Nguyen:
Meta descriptions? There are best practices for that. Title tags? There are best practices for that. Redirects? There are – you guessed it – best practices for that. Welcome to the Search Engine Land podcast, I’m your host George Nguyen. As you’re probably already aware, the internet can be a messy place, SEOs only have so many hours a day and – as IBM SEO specialist Patrick Stox explains – Google may have already accounted for some of the more common lapses in best practices. Knowing which of these items a search engine can

figure out on its own can save you time and allow you to focus on the best practices that will make the most impact. Here's Patrick's Insights session from SMX Advanced, in which he discusses a few of the things Google tries to correct for you.

Patrick Stox:
How's it going? I get to kick off a brand new session type. This should be fun. We're going to talk a little bit about things that Google and, some for Bing, try to correct for you. If you were in the session earlier with Barry [Schwartz] and Detlef [Johnson], they were discussing some of the things that, you know, the web is messy, people make mistakes and it's the same mistakes over and over. And if you're a search engine, what are you going to do? Are you going to throw millions of dollars at a PR campaign to try to get us to convince developers to fix all this stuff? Or are you just going to fix it on your end? And the answer is they fix a ton of stuff on their end.

So the main thing here – I'm here as me. If I say something stupid or wrong, it's me – not IBM.

The importance of technical SEO may diminish over time. I am going to say "may," I'm going to say this with a thousand caveats. The reason being, the more stuff that Google fixes, the more stuff that Bing fixes on their end, the less things we actually have to worry about or get right. So, a better way to say this might be, "it'll change over time" – our job roles will change.

Some of the things: index without being crawled. Everyone knows this. If a page gets linked to Google, sees the links, they're like, here's anchor texts. I know that the page is there. People are linking to it. It's important they index it. Even if we're blocked, you can't actually see what's on that page. They're still going to do it. They're still going to index it.

This is something that happens on both Google and Bing: soft 404s. So what happens with a status code of 200, but there's a message on the page, 200 says okay, there's a message on the page that says something's wrong. Like, this isn't here or whatever. They treat it as a soft 404; this is for Google and Bing. There's literally dozens of different types of messaging where they will look at the page that you just throw a 200 status code on and say, "that's actually a 404 page, and they treat that as a soft 404." They're like, "we know there's not actually anything useful there most of the time." But this happens a lot with JavaScript frameworks because those aren't typically made to fail. You actually have to do some hacky work arounds, like routing, like Detlef talked about, to a 404 page. So, you have thrown in a 200 but they're like page not found. Search engines are like, "no, there's nothing there."

With crawling, crawl delay can be ignored. Google typically will put as much load on the server as your server can handle, up to the point where they get the pages that they want. Pages may be folded together before being crawled. If you have duplicate sections, say like one on a sub domain or like HTTP, HTTPS, they recognize these patterns and say, I only want one version. I want this one source of truth. Consolidate all the signals there. So before, if they've seen it the same way in five different places, then they're going to just treat that as one. They don't even have to crawl the page at that point – they're like, this repeated pattern is always the same.

It kind of works that way with HTTPS, also. This is actually one of the duplicate issues, is that they will typically index HTTPS first over HTTP. So, if you have both and you don't have a canonical – canonical, we could go either way, but typically they're going to choose HTTPS when they can.

302 redirects: I think there's a lot of misunderstanding with SEOs, so I'm actually going to explain how this works. 302s are meant to be temporary, but if you leave them in place long enough, they will become permanent. There'll be treated

exactly like 301s. When the 302 is in place, what happens is if I redirect this page to this page, it actually is like a reverse canonical: all the signals can go back to the original page. But if you leave that for a few weeks, a few months, Google was like, “Nah, that’s really still redirected after all this time. We should be indexing the new page instead.” And then all the signals get consolidated here, instead.

Title tags: Anytime, you know, you don’t write a title tag or it’s not relevant, generic, too long; Google has the option to rewrite this. They’re going to do it a lot, actually. You know, if you just write “Home,” maybe they’re going to add a company name. They’re going to do this for a lot of different reasons, but the main reason I would say is that you know, people were really bad about writing their titles. They were bad about keyword stuffing their titles. And it’s the same with meta descriptions: they’re typically going to pull content from the page. If you don’t write a meta description, they’re going to write one for you. It’s not like, “Hey, that doesn’t exist.”

Lastmod date and site maps – I believe Bing actually ignores this, too. The reason being the sitemap generators, the people making the site maps, this is never ever right. I would say this is one of the things that is probably most wrong, but who cares. They ignore it.

Canonical tags: this is very common. This is like half of my job is trying to figure out how things got consolidated or is something actually a problem. In many cases, the canonical tags will be ignored. Could be other signals in play, like hreflang tags or any number of things. But basically if they think that something is wrong, they’re just going to say, “Nope, canonical is, you know, a suggestion.” It is not a directive. So anytime that they think that the webmaster, the developer, the SEO got it wrong, they’re going to make their best guess at what that should be.

It’s kind of the same with duplicate content. Duplicate content exists on the web. It is everywhere. In Google’s mind, they’re trying to help people by folding the pages together. All these various versions become one. All the signals consolidate to that one page. They’re actually trying to help us by doing that. And they actually do a pretty good job with that.

If you have multiple tags, they’re going to choose the most restrictive. I’ve seen this a thousand times with different CMS systems: in WordPress, you might have your theme adding a tag, plus Yoast adding a tag, plus any number of things can add tags, basically. And usually if there’s five tags that say index and one that’s noindex, they’re going to choose the most restrictive and that’s the noindex.

With links, they’re typically going to ignore them. If you have bad links to your site – I think there was some discussion earlier – are you going to use the disavow file – or this might’ve been last night actually; Barry was talking about this. In general, the answer’s no. If you’re afraid you’re going to have a penalty, maybe, but for the most part you don’t have to worry about the links to your site anymore, which is great.

Then if you’re in local, the NAP listings, a lot of local SEOs we’ll really focus on, like, these all have to be the exact same thing. Well, variations, you know street, spelled out versus “st,” or LLC versus limited liability corporation. There are certain variations where basically they’re going to consolidate. They know that this is another version of this other thing, so they’re going to say it’s the same, it’s fine.

This actually came up earlier too with Barry or Detlef, I can’t remember which, but they were saying that Google only looks at HTTPS in the URL, not whether your certificate is actually valid or not. And that’s 100% true. If you ever crawl a page that has an expired certificate, they go right through.

If you look in search console, all the links consolidate. They follow the redirect that's there even though the user is going to get an error.

And then hreflang, I think again, Barry had mentioned this, this is one of the most complicated things. This is, in my world, the most likely thing that's going to go wrong a million different ways because it really does get complex. With duplicates, they're typically going to show the right one anyway, even if you didn't localize the page at all – like you have 30 versions, all English, as long as the signals are there, it's going to be okay. It's when the tags break and that kind of thing, you might end up with the wrong version showing, cause again, they're folding the pages together; typically, if they're duplicates, and they're trying to show one main version. If everything's right though, they will swap to show the right version for the right person. Within that tag, you know, it's a best practice to use a dash instead of an underscore – doesn't really matter; their crawlers are very lenient. Detlef was talking about like, "oh you got to get their semantic HTML right." Their crawlers have seen this stuff wrong 50 billion different times and honestly they are very lenient on a lot of things.

en-UK instead of en-GB: every hreflang article will tell you this is wrong, but it works. You will never see an error for this. Why? Because UK is not actually a country – it's a reserved code and they've seen it wrong enough that they're like, "Eh, it's fine."

Same with self referencing. You don't actually need that. Same with relative URLs versus absolute. There are best practices basically. But, then there's kind of what works and I think where we have to get as an industry is let's not waste people's time. If Google, if Bing have fixed this on their end, why are we pushing for it? We've got other priorities, other things that we can have done.

They’re even doing this in the browser, now. Most websites do not use lazy loading for their images. Google is going to take that on in the browser and I hope other browsers do this. I think this is the first step. I think they’re going to do a lot more with this, probably like preload directives and a bunch of things, but they’re going to, in the browser, take the strain off the server, off the websites, and they’re just going to be lazy loading images across the web. Now, a lot of people are thinking that they need this loading=“lazy” — that’s actually default. If you do nothing, you have lazy loading on your website as of Chrome 75. And that’s about it, thank you.

About The Author

George Nguyen is an Associate Editor at Third Door Media. His background is in content marketing, journalism, and storytelling.
http://feeds.searchengineland.com/~r/searchengineland/~3/NSzWYCdYNNQ/smx-replay-seo-that-google-tries-to-correct-for-you-319091