An Investigation Into Google's Maccabees Update

December brought us the latest piece of algorithm update fun. Google rolled out an update which was quickly named the Maccabees update and the articles began rolling in (SEJ , SER).

The webmaster complaints began to come in thick and fast, and I began my normal plan of action: to sit back, relax, and laugh at all the people who have built bad links, spun out low-quality content, or picked a business model that Google has a grudge against (hello, affiliates).

Then I checked one of my sites and saw I'd been hit by it.

Hmm.

Time to check the obvious

I didn't have access to a lot of sites that were hit by the Maccabees update, but I do have access to a relatively large number of sites, allowing me to try to identify some patterns and work out what was going on. Full disclaimer: This is a relatively large investigation of a single site; it might not generalize out to your own site.

My first point of call was to verify that there weren't any really obvious issues, the kind which Google hasn't looked kindly on in the past. This isn't any sort of official list; it's more of an internal set of things that I go and check when things go wrong, and badly.

Dodgy links & thin content

I know the site well, so I could rule out dodgy links and serious thin content problems pretty quickly.

(For those of you who'd like some pointers on the kinds of things to check for, follow this link down to the appendix! There'll be one for each section.)

Index bloat

Index bloat is where a website has managed to accidentally get a large number of non-valuable pages into Google. It can be sign of crawling issues, cannabalization issues, or thin content problems.

Did I call the thin content problem too soon? I did actually have some pretty severe index bloat. The site which had been hit worst by this had the following indexed URLs graph:



However, I'd actually seen that step function-esque index bloat on a couple other client sites, who hadn't been hit by this update.





In both cases, we'd spent a reasonable amount of time trying to work out why this had happened and where it was happening, but after a lot of log file analysis and Google *site:* searches, nothing insightful came out of it.

The best guess we ended up with was that Google had changed how they measured indexed URLs. Perhaps it now includes URLs with a non-200 status until they stop checking them? Perhaps it now includes images and other static files, and wasn't counting them previously?

I haven't seen any evidence that it's related to m. URLs or actual index bloat – I'm interested to hear people's experiences, but in this case I chalked it up as not relevant.

Appendix help link

Poor user experience/slow site

Nope, not the case either. Could it be faster or more user-friendly? Absolutely. Most sites can, but I'd still rate the site as good.

Appendix help link

Overbearing ads or monetization?

Nope, no ads at all.

Appendix help link

The immediate sanity checklist turned up nothing useful, so where to turn next for clues?

Internet theories

Time to plow through various theories on the Internet:

1. The Maccabees update is mobile-first related
    - Nope, nothing here; it's a mobile-friendly responsive site. (Both of these first points are summarized here.)
2. E-commerce/affiliate related
    - I've seen this one batted around as well, but neither applied in this case, as the site was neither.
3. Sites targeting keyword permutations
    - I saw this one from Barry Schwartz; this is the one which comes closest to applying. The site didn't have a vast number of combination landing pages (for example, one for every single combination of dress size and color), but it does have a lot of user-generated content.

Nothing conclusive here either; time to look at some more data.

Working through Search Console data

We've been storing all our search console data in Google's cloud-based data analytics tool BigQuery for some time, which gives me the luxury of immediately being able to pull out a table and see all the keywords which have dropped.

There were a couple keyword permutations/themes which were particularly badly hit, and I started digging into them. One of the joys of having all the data in a table is that you can do things like plot the rank of each page that ranks for a single keyword over time.

And this finally got me something useful.

The yellow line is the page I want to rank and the page which I've seen the best user results from (i.e. lower bounce rates, more pages per session, etc.):



Another example: again, the yellow line represents the page that should be ranking correctly.



In all the cases I found, my primary landing page – which had previously ranked consistently – was now being cannabalized by articles I'd written on the same topic or by user-generated content.

Are you sure it's a Google update?

You can never be 100% sure, but I haven't made any changes to this area for several months, so I wouldn't expect it to be due to recent changes, or delayed changes coming through. The

site had recently migrated to HTTPS, but saw no traffic fluctuations around that time.

Currently, I don't have anything else to attribute this to but the update.

How am I trying to fix this?

The ideal fix would be the one that gets me all my traffic back. But that's a little more subjective than "I want the correct page to rank for the correct keyword," so instead that's what I'm aiming for here.

And of course the crucial word in all this is "trying"; I've only started making these changes recently, and the jury is still out on if any of it will work.

No-indexing the user generated content

This one seems like a bit of no-brainer. They bring an incredibly small percentage of traffic anyway, which then performs worse than if users land on a proper landing page.

I liked having them indexed because they would occasionally start ranking for some keyword ideas I'd never have tried by myself, which I could then migrate to the landing pages. But this was a relatively low occurrence and on-balance perhaps not worth doing any more, if I'm going to suffer cannabalization on my main pages.

Making better use of the Schema.org "About" property

I've been waiting a while for a compelling place to give this idea a shot.

Broadly, you can sum it up as using the About property pointing back to multiple authoritative sources (like

Wikidata, Wikipedia, Dbpedia, etc.) in order to help Google better understand your content.

For example, you might add the following JSON to an article an about Donald Trump's inauguration.

[{ "@type": "Person", "name": "President-elect Donald Trump", "sameAs": ["https://en.wikipedia.org/wiki\Donald_Trump", "http://dbpedia.org/page/Donald_Trump", "https://www.wikidata.org/wiki/Q22686"] }, { "@type": "Thing", "name": "US", "sameAs": ["https://en.wikipedia.org/wiki/United_States", "http://dbpedia.org/page/United_States", "https://www.wikidata.org/wiki/Q30"] }, { "@type": "Thing", "name": "Inauguration Day", "sameAs": ["https://en.wikipedia.org/wiki/United_States_presidential_inauguration", "http://dbpedia.org/page/United_States_presidential_inauguration", "https://www.wikidata.org/wiki/Q263233"] }]

The articles I've been having rank are often specific sub-articles about the larger topic, perhaps explicitly explaining them, which might help Google find better places to use them.

You should absolutely go and read this article/presentation by Jarno Van Driel, which is where I took this idea from.

Combining informational and transactional intents

Not quite sure how I feel about this one. I've seen a lot of it, usually where there exist two terms, one more transactional and one more informational. A site will put a large guide on the transactional page (often a category page) and then attempt to grab both at once.

This is where the lines started to blur. I had previously been on the side of having two pages, one to target the transactional and another to target the informational.

Currently beginning to consider whether or not this is the correct way to do it. I'll probably try this again in a couple places and see how it plays out.

Final thoughts

I only got any insight into this problem because of storing Search Console data. I would absolutely recommend storing your Search Console data, so you can do this kind of investigation in the future. Currently I'd recommend paginating the API to get this data; it's not perfect, but avoids many other difficulties. You can find a script to do that here (a fork of the previous Search Console script I've talked about) which I then use to dump into BigQuery. You should also check out Paul Shapiro and JR Oakes, who have both provided solutions that go a step further and also do the database saving.

My best guess at the moment for the Maccabees update is there has been some sort of weighting change which now values relevancy more highly and tests more pages which are possibly topically relevant. These new tested pages were notably less strong and seemed to perform as you would expect (less well), which seems to have led to my traffic drop.

Of course, this analysis is currently based off of a single site, so that conclusion might only apply to my site or not at all if there are multiple effects happening and I'm only seeing one of them.

Has anyone seen anything similar or done any deep diving into where this has happened on their site?

Appendix

Spotting thin content & dodgy links

For those of you who are looking at new sites, there are some quick ways to dig into this.

For dodgy links:

- Take a look at something like Searchmetrics/SEMRush and see if they've had any previous penguin drops.
- Take a look into tools Majestic and Ahrefs. You can often get this free, Majestic will give you all the links for your domain for example if you verify.

For spotting thin content:

- Run a crawl
  - Take a look at anything with a short word count; let's arbitrarily say less than 400 words.
  - Look for heavy repetition in titles or meta descriptions.
  - Use the tree view (that you can find on Screaming Frog, for example) and drill down into where it has found everything. This will quickly let you see if there are pages where you don't expect there to be any.
  - See if the number of URLs found is notably different to the indexed URL report.
- Soon you will be able to take a look at Google's new index coverage report. (AJ Kohn has a nice writeup here).
- Browse around with an SEO chrome plugin that will show indexation. (SEO Meta in 1 Click is helpful, I wrote Traffic Light SEO for this, doesn't really matter what you use though.)

Index bloat

The only real place to spot index bloat is the indexed URLs report in Search Console. Debugging it however is hard, I

would recommend a combination of log files, “site:” searches in Google, and sitemaps when attempting to diagnose this.

If you can get them, the log files will usually be the most insightful.

Poor user experience/slow site

This is a hard one to judge. Virtually every site has things you can class as a poor user experience.

If you don’t have access to any user research on the brand, I will go off my gut combined with a quick scan to compare to some competitors. I’m not looking for a perfect experience or anywhere close, I just want to not hate trying to use the website on the main templates which are exposed to search.

For speed, I tend to use WebPageTest as a super general rule of thumb. If the site loads below 3 seconds, I’m not worried; 3–6 I’m a little bit more nervous; anything over that, I’d take as being pretty bad.

I realize that’s not the most specific section and a lot of these checks do come from experience above everything else.

Overbearing ads or monetization?

Speaking of poor user experience, the most obvious one is to switch off whatever ad-block you’re running (or if it’s built into your browser, to switch to one without that feature) and try to use the site without it. For many sites, it will be clear cut. When it’s not, I’ll go off and seek other specific examples.

https://moz.com/blog/an-investigation-into-googles-maccabees-update?utm_medium=feed&utm_source=feedpress.me&utm_campaign=Feed%3A+mozblog