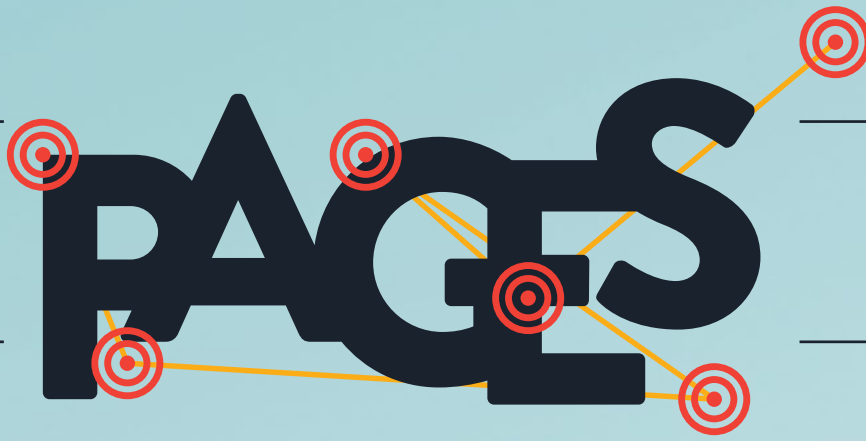


<!doctype mag>

VOLUME 1
ISSUE 3
2018



CLEAN
PROVEN
SEO

<title>

A N S E O M A G A Z I N E

</title>

<body>

<H1> THE TECHNICAL SEO ISSUE </H1>

<meta> July 5, 2018 - This issue of *PAGES* explores technical **SEO**: search engine optimization techniques concerning the **technical** elements of a website, which improve the site's performance in **search**. </meta>

<content> </content>

GOING META ON YOUR SEO

Dr. Peter Meyers
Marketing Scientist, Moz

DIAGNOSING TRAFFIC DROPS

Marie Haynes
Marie Haynes Consulting, Owner

STRUCTURED DATA

Alan Ibbett
Chief Information Officer, Diocese of Wollongong

AN ANSWER YOU CAN'T IGNORE

Charlie Williams
SEO Strategist, Screaming Frog

RACE TO THE TOP

Wes Marsh
Director of Digital Marketing, SoloDev

SOLVING CRAWLER ISSUES

Nicholas Chimonas
Director of SEO, WTP Inc.

</body>

</mag>



pages_cover_image

We

1101001011001
01010010010110100010101
0111010101101100101101010011
01010110010100100011110001101110
01001100011010010110000101101100100
00001000000111000101110101011000011101
Pages_in_Index = 219B110010101101110011110
01110101011011000111011001101001011011
01000110100101100001011011010010000001
1110101110010000001100101010110001000
000111001101100101011011010111000001
1000101110101011011110111001000100
10010000001110110011010001011001
Content_Explorer_Index = 944M pages1010101110011
010010111010001100001011011
11000100110100100100000011
010010111000101110101011
1100101011000110110011
Hardware_servers = 2.8K0001100
010111001101110101
1011011010001101
CPU_cores = 65K000110110
011100110010
0001101110
01000110
000011
0110
10
1

SEO TOOLS POWERED
BY SERIOUSLY BIG DATA.
TRY US AT [AHREFS.COM](https://ahrefs.com)

0100100101101
11010010010100101010010
000011101010111001 **Unique_Domains = 186.8M**
10100101100000010000100000011000
11010100100100001110110011001010110
01011011010010000001100101001010011101
01000111010101101101001000010101000111
100110000101001001 **External_Backlinks = 2.3T**
10111001101111011011100010000111000100
0010100001101110100010111001100100010
100110 **Internal_Backlinks = 16.6T** 1011
0000110110001011101010110010000110
01001000000111010001100101010001
001011001100001011000100111101
100111100100000011011100110
101011100110 **Keywords_Index = 5.9B Keywords**
011011001010110111000101
0010000001100101010111
10101110101110010000
000110110101100011
1101000011101101
00110000010101
001000111010
1000100111
00010101
101100
0110
11
0

big
data.

ahrefs





WELCOME.





EDITOR IN CHIEF: JOE OLIVER

ART AND DESIGN: A.J. OGDEN

MANAGING EDITOR: SLOAN ROSEBERRY

MARKETING AND ADVERTISING: JAN SCHMIDT

PHOTOGRAPHY: UNSPLASH.COM

ADVERTISING DESIGN: MOLLY SCHMIDT

PRODUCTION: PAGE ONE POWER

PUBLISHING & PRINTING: EVERMORE PRINTS LLC

MAIN OFFICES: 3100 N LAKEHARBOR LN STE 100 BOISE, ID 83703 208.229.7046

FOR ADVERTISING INQUIRIES AND PLACEMENT: JAN@pagesSEOmagazine.com

COPYRIGHT 2018 BY PAGE ONE POWER

 [@PAGESSEOMAG](https://twitter.com/PAGESSEOMAG)

 [FACEBOOK.COM/PAGESSEOMAGAZINE](https://www.facebook.com/PAGESSEOMAGAZINE)

 [@PAGESSEOMAG](https://www.instagram.com/PAGESSEOMAG)

 WWW.PAGESSEOMAGAZINE.COM



FEATURED

articles



GOING META ON YOUR SEO

Dr. Peter Meyers
Marketing Scientist, Moz

P. 18



AN ANSWER YOU CAN'T IGNORE

Next Level Optimization
Charlie Williams
SEO Strategist, Screaming Frog

P. 42



DIAGNOSING TRAFFIC DROPS

Marie Haynes
Owner, Marie Haynes Consulting, Inc.

P. 24



RACE TO THE TOP

Quick Tips For A Faster Page
Wes Marsh
Director of Digital Marketing, SoloDev

P. 48



STRUCTURED DATA

Alan Ibbett
Chief Information Officer,
Catholic Education Office in the Diocese of Wollongong

P. 32



SOLVING CRAWLER ISSUES

Nicholas Chimonas
Director of SEO, WTP Inc.

P. 58

more CONTENTS

10 LETTER FROM THE EDITOR

14 READER VOICES

15 TIME KILLER

16 DIY TECHNICAL AUDIT

28 INDUSTRY BOSS: MARIE HAYNES

52 SEEING DOUBLE: DEALING WITH DUPLICATE CONTENT

64 THE INDEX LANDSCAPE

65 TECHNICAL SEO GLOSSARY





MEASURE TWICE, CUT ONCE

Have you ever heard how over 90% of an iceberg's volume and mass is underwater?

What we see is only a tiny portion of the ice as a whole. The same goes for the construction of a building, or the development of a city or subdivision. What you see are the cosmetic elements that make the structure appealing and inviting, but there is much more happening behind the scenes.

Take a house, for example. Behind the walls and under the floors are a maze of wires, plumbing, insulation, and the actual wood that makes the frame of the house. Under the house is a foundation where cement and metal rods hold the whole structure together. Feeding the house is the city plumbing and electrical. All of this is hidden from view, but essential for the house to function correctly.

The house then turns into a home when we add our furniture, paint, appliances, pictures, and electronics. The parts we enjoy, interact with, and that make a house a home can only be appreciated when the unseen elements are functioning correctly.

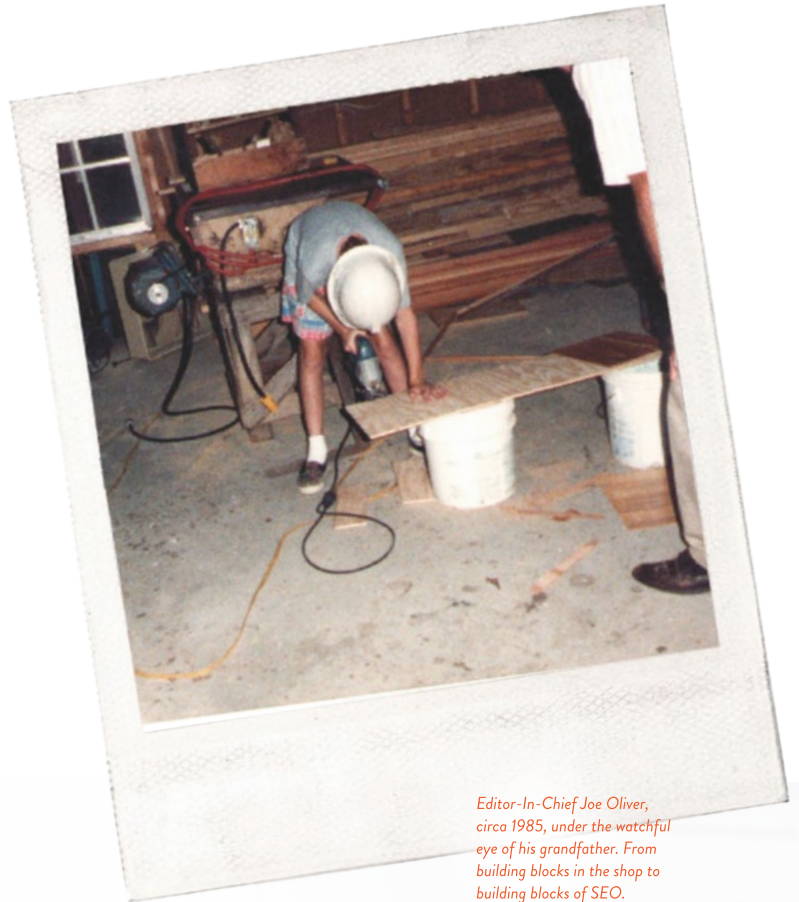
Websites are no different: the building blocks that make a website perform well in search are the things most people would never know even existed. The end user doesn't know what is happening behind the beautiful images, creative design, or functionality — nor should they. Most people don't think about why, when turning the light switch on, the light appears. They're not thinking about the inner workings of electrical power flowing through the walls, ceiling, or floors of the house, providing currents at the right voltage to power the light bulb giving light to a room.

Technical SEOs cannot be successful in search without this knowledge. We have to ensure that the elements that make a house are in place so that people can find and enjoy the home.

Our issue this quarter gathers insights from the brightest in the SEO industry regarding technical search engine optimization. The concepts we'll explore may not be ones you're already familiar with, but trust me when I say, ignore technical SEO and people won't be able to find your house — let alone make it a home.

Thank you once again for reading.

Joe Oliver



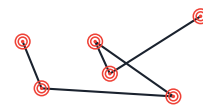
Editor-In-Chief Joe Oliver, circa 1985, under the watchful eye of his grandfather. From building blocks in the shop to building blocks of SEO.

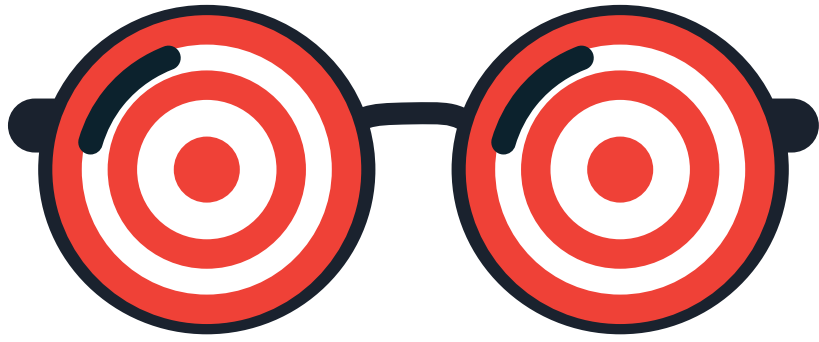
**LETTER
FROM
THE
EDITOR**

VOL. 1 ISSUE 3

VOLUME 1
ISSUE 3
2018

THE TECHNICAL SEO ISSUE





Your dose of old school



brought to you by:



PAGES



A N S E O M A G A Z I N E

www.pagesseomagazine.com



pagesSEomag



PAGES SEO Magazine



PAGES SEO Magazine



pagesSEomag

READER VOICES

ON PAGES MAGAZINE



Got something to say about PAGES? Let your voice be heard.

We love hearing from you! Share your thoughts about PAGES on social media using the hashtag below, or send us an email at: info@pagesSEOMagazine.com

#PAGESSEOMAGAZINE



Sunday reading. Getting ready for the new week. @pageonepower @joeOliver #SEO #sundaymotivation #pagesmagazine @STATICSETH



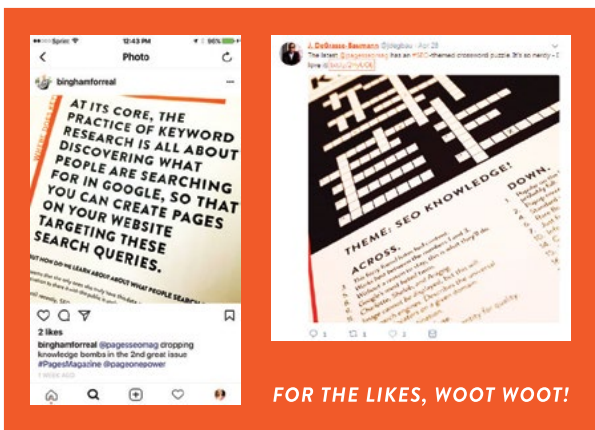
Just got the new PAGES Magazine. Great SEO read. Check out the packaging! Thanks so much for sending, can't wait to read. @VINCETHENERO



Nicely done. Well written. Quality paper. Oversized. Loved it. Most go in the recycle bin when I'm done, but not this one. @JOEWATERS



I just got a copy of @pageonepower's new magazine, PAGES. It's really sleek! Thanks guys. @fighto PAUL SHAPIRO



FOR THE LIKES, WOOT WOOT!

CORRECTIONS

To any of our readers who were finding themselves spending a lot of time figuring out #3 across in last issue's brain teaser: we're very sorry we caused your brain a lot of undue pain. Thank you Laura Gillespie for the catch!



Laura Gillespie @lauragi23 · May 4 Got my first edition of Pages from @pageonepower - sat down to do the crossword and was stumped by 3 across. Checked the answers in the end to find it has been misprinted. 2 squares down instead of 3. Talk about a time killer 🤔

I'M STILL A FAN OF PAPER, AND I LOVE THAT PAGES IS A THING.

@BEAUPEDRAZA

Thanks

TWEETS A MILLION!



THOMAS'S

IPS
S



Thomas Swisher is Page One Power's resident Web Analyst, SEO Strategist, and Mad Data Scientist extraordinaire.

HERE'S THOMAS'S TAKE ON A FAVORITE SEO EXTENSION FROM HIS OWN TOOLBAR:



SEO Quake

SEO Quake is a plugin that provides data about a wide variety of SEO metrics. Along with at-a-glance page overviews, this extension has several very useful SERP analysis features, making this suite of tools quite impressive for the price (free). Here are some of my favorite ways to use the tool:

THE SEOBAR

SEO Quake primarily interfaces with a toolbar, the SEObar. This toolbar gives you a summary of any webpage's optimization based on 26 ranking parameters. The toolbar is customizable, making it easy to focus on the metrics most important to you. SEO Quake also provides onpage keyword density reports.

OVERLAY

Additionally, SEO Quake has some helpful SERP overlay features. The tool makes it easy to download SERPs for further vetting, and provides a customizable SERP report that gives statistics about the sites that result from your query. It also gives you the option to highlight any domain you'd like to see in the SERPs when present.

TIME KILLER



AKA: BRAINBUSTER

SEO WORD SEARCH
GOOD LUCK!

Answers arranged in all directions

U F Z C N J N M H K T N E T N O C T V K
 Y F Q M C L R R F Q J M I I C C R I M N
 Q E R V H O H A X N J M V T T O B T G C
 T S I U W Y E F N Q S U K A X O P L E K
 M I H L P H P S T K B G Q C A C C E U H
 R T B R U C K E L I I B R L N C E T S W
 C R T Q B J G A R A D N A R A U S A S T
 G E J K L V R Y W L C U G R L R S G E E
 G P S M L U W T B D I I A J Y R Z L N H
 N X I E B O N I F P F N N J T A I C I Z
 I E M T I B N R U A O F K H I N T O H D
 K H M A G S P O N M F Y D Z C C D M T B
 C A Q W V S T H I E Z A H T S E K P R O
 A U E U C Z N T M T F P N H U E T E O B
 R U G R I E E U T I A E U S L R K T W N
 T P A W X F T A T S T N A Y F F P I T I
 U P P S F Q N I I A L M I A W P F T S A
 E L A C X Q I T L I P D S G G V J I U M
 H D T H I N B B O Q B A S Z A A V O R O
 E E V G D Y Y F I N I M M X D P K N T D

**ANALYTICS
COMPETITION
DOMAIN
INTENT
MINIFY
SCRAPE
TITLE TAG**

**AUDIT
CONTENT
EXPERTISE
LATENT
PAGINATION
SITEMAP
TRACKING**

**AUTHORITY
CO-OCCURRANCE
HYPERLINK
META
RANKING
THIN
TRUSTWORTHINESS**

CAN YOU FIND THE SECRET BONUS WORD?

It's hidden in the word search! When you find it, let us know on Twitter using the tag #pagesSEOMagazine. The first reader to find the bonus word will receive a special surprise!

While SEO Quake isn't yet available with the new version of Google Search Console (GSC), integration with GSC allows you to view significant updates when reviewing important events (like clicks and impressions), and review the Domain Score and Trust Score of a page when reviewing links in GSC.

The broad functionality of this tool makes it one of the best free SEO tools available if you're looking for easy access to analytical data on any webpage.

SEO Quake is compatible with Mozilla Firefox, Google Chrome, and Opera. ©



DO-IT-YOURSELF

TECHNICAL AUDIT

THIS QUICK & DIRTY TECHNICAL AUDIT CAN HELP YOU IDENTIFY ISSUES ON YOUR SITE



START

CHECK THE INDEX

How's it look?

Looks good!

Something's not quite right...

HOW TO CHECK THE INDEX?

Try searching site:domain.com in the search engine, or use a tool like WebSite Auditor.

GOOGLE ANALYTICS all set up?

GOOGLE ANALYTICS tracks and reports website traffic. It does this using a snippet of code that's added to each page of your website. This code collects anonymized visitor data that can help you understand where your traffic is coming from. This information can help guide your marketing decisions and give you a look at how your current methods are performing. GA is also integrated with Google advertising services AdWords and AdSense.

Double check these essentials:

1. Tracking code is on every page and event tracking is set up.
2. There's only 1 instance of a given Google Analytics property on a page.
3. Onsite searches are tracked.
4. Internal IP addresses are excluded.

I think so...

Yes!

Review your disallowed pages in robots.txt to ensure all important pages get added to the list. You can also disallow pages at the page level using the "noindex" tag in HTML pages or the HTTP header.

And connected to GOOGLE SEARCH CONSOLE?

You know it!

This one gets you under the hood of the search side of your site, and it facilitates communication between you and Google. You want to get your hands on the info here - it's good stuff!



Not yet... :(

Are you mobile-friendly?

We could be a little closer...

Best buds!

It's time to put down the proverbial flip phone. Head over to Google's mobile-friendly test tool and give it a whirl - your mobile website will soon be the "default" version of your site, as far as Google is concerned. You'll need to check your analytics to find out if the majority of your traffic really comes from mobile and plan accordingly. Mobile sites need regular technical SEO audits, just like their desktop counterparts.

Great, thanks for asking!

How are those internal links?

Internal links are important: a logical structure helps with crawlability, UX, and distributing link equity.

It's time for some housekeeping...

GOOGLE SEARCH CONSOLE (GSC) is your most direct look at your site's relationship with Google. Using GSC, you can show Google which of your pages to index & show in results. GSC also allows you to identify and fix errors on your site, and is a portal to a lot of very useful information about your site's visitors. You can also manage your Accelerated Mobile Pages (AMP) and connect GSC with Google Analytics to improve your targeting.

CONTENT CONSIDERATIONS

CREATIVE STUFF:

Content needs to be unique, informative, & keyword-targeted in order to rank. Keyword stuffing - loading copy with a target keyword, even if it made little sense to actual readers - was a tactic commonly used to help a site rank in the past, but search engines got smart and now reward content that web users actually engage with.

TECHNICAL STUFF:

Title tags: are they unique & 55-60 pixels?

H1 tag: does it contain the primary keyword phrase?

Images: do the file names and alt text contain the primary keyword?

URLs: should be short, clean, optimized, and descriptive.

CRAWL BUDGET is the number of your site's pages that search engine will crawl on a given day. Search engines set the budget and there's a correlation between the number of backlinks a site has and how many web crawlers or "spiders" visit on a given day.

Take a look at your SITEMAP.

xml SITEMAP XML sitemaps ensure spiders can easily find and crawl all of your site's essential pages. They give you the opportunity to create a clear guide for Google about which of your pages you believe are important and deserve to rank.

Get like Ptolemy and keep your sitemap updated, clean, and short for effective crawls. Make sure it's listed in robots.txt and has been submitted to GSC.

It's looking lovely!

Exceptionally healthy!

What about on-page CONTENT?

It's a little skinny...

Time to start bulking! Get some solid content on your site. You need more real stuff than ads, and enough words on the page for people and search engines to know what it's about - that's usually a few hundred words.

Are you up to speed?

Slow and not-so-steady...

How did my dust taste?

Page speed is more important than ever with mobile users and it's another major ranking factor. Check out the recommendations Google's PageSpeed insights tool has for your site to get things moving. If you want to go a little further, try out the GTmetrix tool for a real-time speed analysis, tracking, and competitor analytics.

Quick tip: Images can be a big issue here.

Is your CRAWL BUDGET optimized?

Not yet...

Darn tootin' it is!

Here's 4 things that can help:

1. Remove duplicate pages.
2. Prevent pages without SEO value from being indexed.
3. Grow your backlinks.
4. Repair broken links.

Are you international?

Mr. Worldwide, yo...

Domestic only, locals only!

Another quick 4 things to check:

1. Are the currencies correct for the target country?
2. Is the URL structure in the native language?
3. International versions of sites in the same language can cause duplicate content issues - be sure to update the content so it's unique to each version of the site.
4. Did you enable geo-targeting in GSC?

Sure are!

Are resources on your site like CSS & JavaScript indexed?

That would be a no...

Looks matter. Google needs to see how your site is intended to look to properly index and rank it. Get these resources added to help the crawlers see your site for what it really is.

GREAT JOB

YOUR BASIC TECHNICAL AUDIT IS COMPLETE!



GOING META ON YOUR SEO

DR. PETER J. MEYERS

MARKETING SCIENTIST, MOZ

Whether you've been an SEO for two weeks or two decades, you're inevitably going to stumble into arguments about meta tags. It's sometimes hard to separate myth from history, and the role of meta tags in SEO has evolved a great deal over time, but the basics are still critical and don't need to be complicated.

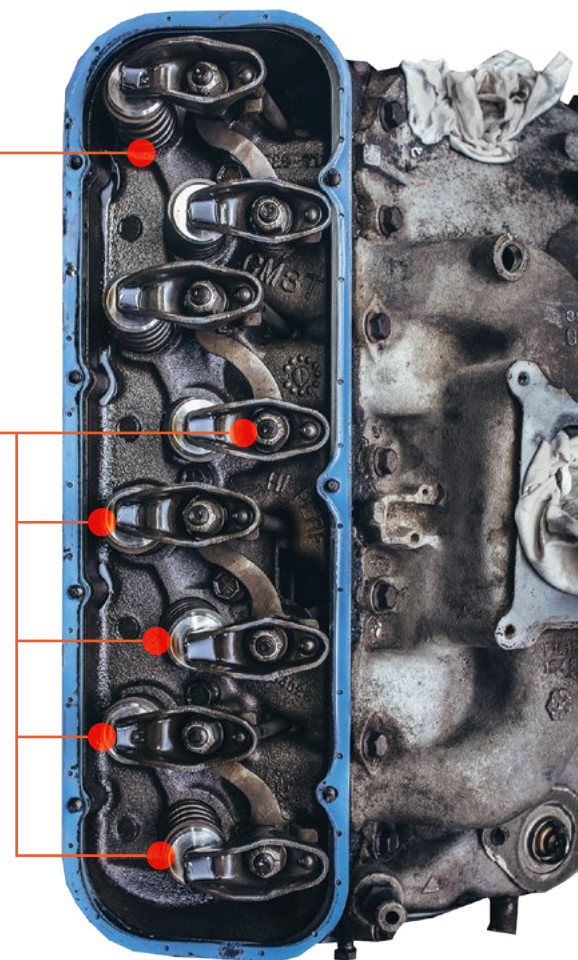
Put simply, meta tags are data that help search engines understand your webpage but aren't part of the regular content. There are a number of advanced meta tags, but let's focus on the three with the most SEO history: Meta keywords, Meta description, and Title tag, in reverse importance...

META KEYWORDS

The Meta keywords tag is a comma-delimited list of keywords (or topics) that your webpage is about. There was a time when this tag helped search engines decide where and when your page should rank.

That time was twenty years ago.

By the early 2000s, search engines realized that Meta keywords were too easy to game, and we have good reason to believe that many even viewed them as a negative ranking factor (i.e. they treated those keywords with suspicion). In 2018, the Meta keywords tag is unlikely to help or harm you in most situations, and it isn't worth much time or effort.



META DESCRIPTION

Your Meta description is meant to be a human-readable summary of your webpage. This summary is often (note: not always) used by search engines and some social networks as the description or “snippet” that they display.

This summary is a vital part of the first impression that will determine if someone will click on your search result.

HERE ARE THREE TIPS FOR WRITING GOOD META DESCRIPTIONS:

Write for Humans

Too many people still treat their Meta descriptions as a dumping ground for their SEO keywords. Don't do this. Your Meta description isn't a significant ranking factor in 2018. It is, however, very important in determining whether someone clicks on and engages with your site. Like good ad copy, your Meta description should grab a searcher's interest and leave them wanting more.

Mind Your Length

Google currently limits display snippets to about 155 characters, after which they add an ellipsis (...). Try to keep your Meta description under this length, or make sure that the most important parts of your description happen before the cutoff.

Keep it Relevant

While you want to attract clicks, make sure that your description is relevant to your page's content. Tricking searchers into a click leads to high bounce rates and no long-term value. In many cases, if your description isn't relevant to the search or your content, search engines will simply overrule you and rewrite what they display.





TITLE TAG

Your Title tag not only serves as the page title displayed by web browsers, but is often used by search engines and social media sites as the headline for your page. Keywords in your Title tag do still seem to impact rankings in 2018, but like your Meta descriptions, I'd encourage you to write Title tags for humans and treat them as your first impression to a world that may not know your brand or your content.

HERE ARE THREE TIPS FOR WRITING HIGH-IMPACT TITLE TAGS:

● Think Like a Reporter

Think of your Title tag like a headline, and don't bury the lede. Especially on the web, attention spans are short, and you have to help people quickly understand what your page is about and why they should care.

● Focus on the Unique

Front-load the most unique part of your title. If you have hundreds or thousands of products pages, and you start every Title tag with your brand name, product category, and subcategory, people will give up before they know which product your page is about.

● Be Succinct

In addition to the problem of short attention spans, Google only displays about 50-60 characters of your Title tag. Get to the point quickly, or it's likely to get stranded behind the "..." and never be heard from again.

WRITE RESPONSIBLY

Meta tags are your webpage's face to the search world. If you stuff them with irrelevant nonsense, you're not going to achieve anything in 2018 but irritating potential customers. Write like a human, be relevant, and take the time to craft Meta descriptions and Title tags that represent your value proposition and leave people wanting more. 🌐

TECHNICAL SEO HAS ALWAYS BEEN FOUNDATIONAL TO SEARCH MARKETING SUCCESS.

The advancements made in machine learning over the past number of years, and Google and Bing's algorithms aligning toward this evolution in AI, serve to underscore the critical importance of technical SEO acumen. Content may be king, but a slow, poorly architected website with the world's greatest content will face an uphill battle trying to surface in the SERPs.

-MICHELLE ROBBINS

*Editor-In-Chief,
Search Engine Land*



- MARIE HAYNES -

OWNER, MARIE HAYNES CONSULTING INC.

DIAGNOSING

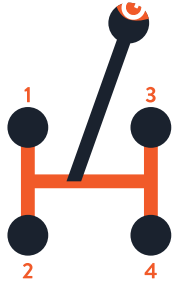
T R A F F I C



A drop in visitors to your site can be one of the first indicators that something is awry with your optimization. From technical errors to regular updates to the Google algorithm, there can be many possible reasons behind a traffic loss. By discovering the cause of the drop, you can begin to optimize and regain your losses.

HAS YOUR WEBSITE SEEN A DECLINE IN TRAFFIC? HERE ARE SEVERAL OF THE MOST COMMON CULPRITS >>>

GOOGLE MANUAL ACTION



This type of penalty is less common than it used to be a few years ago. Still, if your site has seen declines in traffic, one of the first places to look for a problem is in Google Search Console > Search Traffic > Manual Actions.

RECENT SITE CHANGES

If a site's traffic has dropped shortly after making major changes, it's important to see if those changes had a potential impact on SEO. Have URLs been changed? Have title tags been changed? Have there been major changes in content or design? Has the Google Analytics tracking code been left off of new pages?

In order to determine whether significant changes have been made, you can look at previous versions of the site on archive.org.



GOOGLE ALGORITHM CHANGE

Google often makes significant changes to their algorithm that can negatively impact many sites. Google's Panda algorithm, first released in 2011, had a drastic impact on the traffic of many low-quality websites. Google's Penguin algorithm, which launched in 2012, initially penalized websites that were deemed to have a large number of backlinks created solely for SEO purposes.

Now, both Panda and Penguin are a part of Google's main algorithms. In the past, when these algorithms would run only several times per year, it was easy for us to diagnose whether a site had been hit by a Panda or Penguin update.

These days, we tend not to see obvious Panda and Penguin updates. Rather, most of the impactful updates that Google runs are what we call "core quality updates." In some cases, we can document known or suspected updates to the core quality algorithm.

In order to determine if your website saw drops because of an update to Google's quality algorithms, you will need to look at Google organic traffic. In Google Analytics, you can do this by going to Acquisition > Source/Medium and selecting Google/organic. You'll then be able to see if significant, sustained traffic drops have happened that coincide with known or suspected algorithm updates.

Below are two resources you can use to stay in the loop with the latest changes to the algorithm.

LOSS OF A HONEYMOON BOOST

It is not uncommon for a new site to rank quite well for several months, then start to see declines as Google determines their true rankings. If your site is seeing traffic drops and is less than six months old, there is a good chance that you had received an extra boost after first appearing on the scene and have now lost the boost.

MIGRATION TO HTTPS



Migration from HTTP to HTTPS is equivalent to a domain name change. Even if done correctly, many sites will see a temporary dip in rankings after an HTTPS migration. Most sites recover within a few weeks, or, if it is a large site, it can take a few months for recovery to fully happen.

If you see a traffic drop after an HTTPS migration, make sure that all URLs are properly redirected, that the HTTP version properly redirects to HTTPS, that internal links all point to the HTTPS version, and also, that the site's disavow file (if there is one) is uploaded to the HTTPS version in Google Search Console.



GOOGLE ALGORITHM UPDATE HISTORY



pagesseomag.com/go2MarieHaynes



GOOGLE ALGORITHM CHANGE HISTORY



pagesseomag.com/go2moz



LOSING GROUND TO COMPETITORS

This is by far the most common reason for a site to see traffic drops.

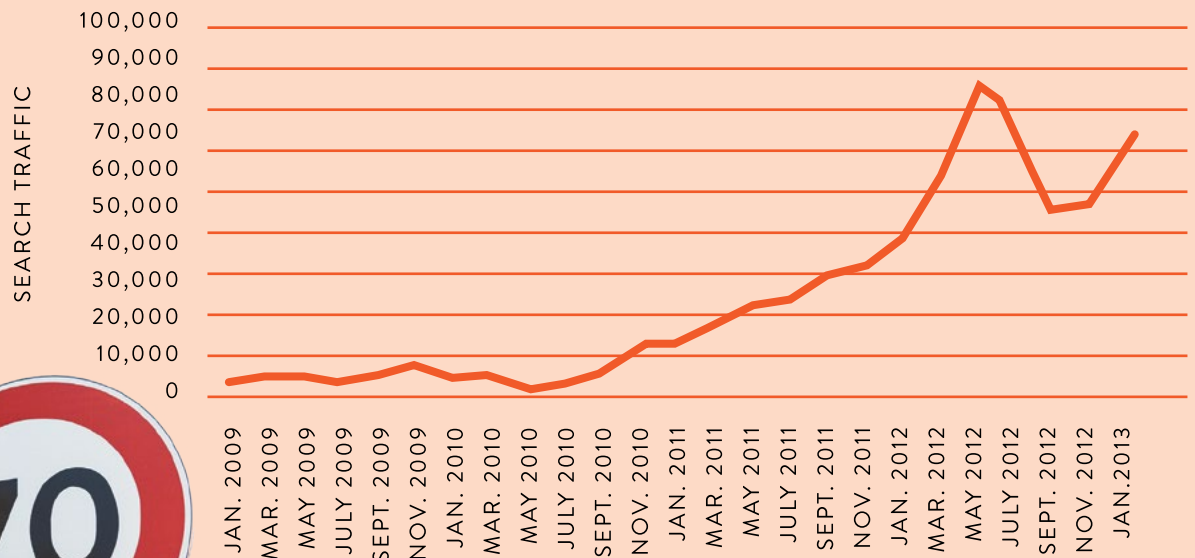
Google is getting better and better at determining which websites are truly doing the best job at helping searchers. If you've lost rankings, it's highly recommended to have several people who are not emotionally connected to your business review both your site and your competitors' sites and give their opinion on whether your site truly is the best result for those searching for your top keywords.

In the past, we could take a mediocre site and, with enough link building and on-page tweaking, make it rank well. But now, the key to being ranked #1 is to truly be the best option for searchers. If you have relied on tricks and loopholes to rank in the past, there is a good chance that Google is no longer rewarding those tactics. 🎯

USING UNNATURAL LINKS



The new Penguin algorithm is really good at just ignoring links that Google would consider unnatural. However, there is a strong possibility that there are other algorithms that can algorithmically demote a website if it is detected that the site has been creating a large number of manipulative links. If you think this is the case for you, diligent use of Google's disavow tool may help.



MARIE

Haynes

Marie is the mind behind Marie Haynes Consulting Inc.

She and her team help businesses perform at their full potential in search.


Marie is a well-known speaker at industry shows like Pubcon and SMX. She's recognized for her thorough coverage of Google's updates, her passion for problem solving, and her active involvement in the search community.

Some of her favorite topics in SEO include penalties, Google algorithm updates, and using the information in Google's Quality Rater's Guidelines to optimize.

You can often find her on Twitter, where she shares her latest insights on Google updates.

 [@Marie_Haynes](https://twitter.com/Marie_Haynes)





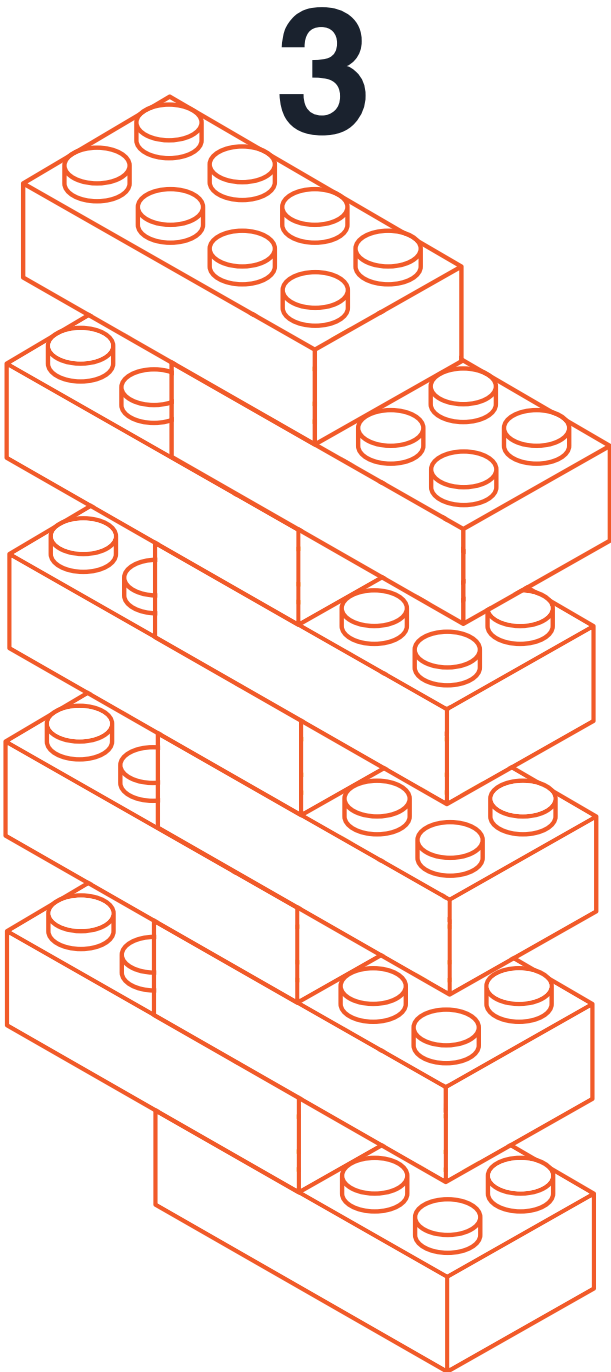
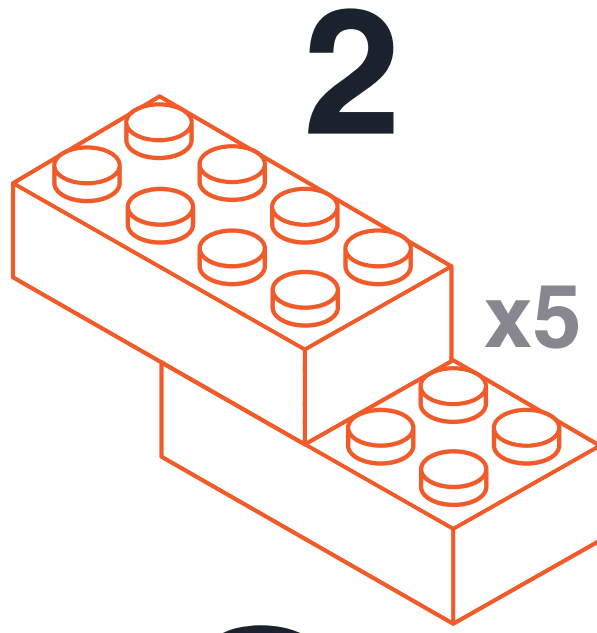
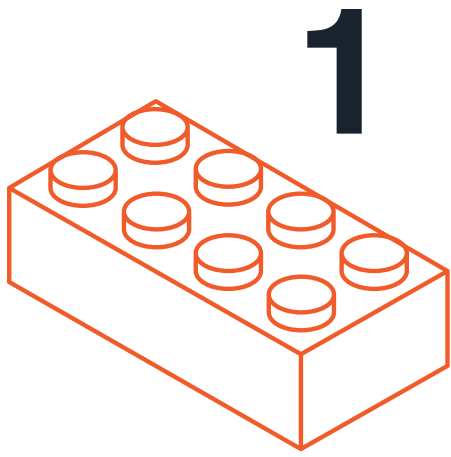
**...WHEN IT COMES
DOWN TO IT, THEY
DON'T WANT "OPTIMIZED"
TEXT OR "OPTIMIZED"
TITLE TAGS. THOSE DO
PROVIDE VALUE, BUT
THOSE ARE JUST SEO
SIGNALS TO THE
BIGGER PICTURE OF
OPTIMIZING THE SITE.**

-STONEY DEGEYTER

Pole Position Marketing

In reference to considering
what it is Google wants in order
to get to the top of Google.





CREATING CONTEXT WITH

STRUCTURED DATA

Alan Ibbett

Chief Information Officer,
Catholic Education Office in the
Diocese of Wollongong

So you have this great website.

The graphics and design are dialed, the HTML is letter-perfect, and you even have web analytics switched on (and you monitor them obsessively). But still, you'd like to see your site bumped up the search results pages.

What else would make it easier for web crawlers to index your site and see what information you have to offer? Turns out there are more tricks in the bag, you just have to reach a bit deeper.

One great technique we can use to help the spiders is called structured data.

This is where art meets engineering.

STRUCTURED DATA EXPLAINED

Structured data is a fancy way of writing information down in an ordered way. That may sound complicated, but it's not, really.

Structured data is behind familiar SERP features, including the Knowledge Graph, rich snippets, and AMP carousels. By properly implementing structured data on your site, you have the opportunity to rank for these SERP features. Structured data also helps crawlers make contextual sense of your site.

For example, consider how many ways there are of presenting the address of your business.

It could be written out as, "The blue coffee van outside the green building on Main Street." That's one way to do it — locals would be able to find your business. However, most of us want to be more specific, so you might give

the address as, "42A Main Street." That's a little better, but there are a few Main Streets around, so then you might add the city, state and, if your coffee is really good, country to the address. Okay, now it's job done.

Think twice: web crawlers are good at picking out details like addresses from common text, but they are not perfect. As it turns out, the design on your site may easily allow a human to pick out the address from the site layout, but an artificially intelligent robot may have more trouble. Your great-looking website actually hides information from the web crawlers who index and set the search results — gosh.

If only we were able to ensure this information could be put on the site in a way that was unambiguous to the web crawler, without inhibiting creative freedom.

Enter, structured data.

1

THE ORIGIN OF STRUCTURED DATA

Implied in its very name, structured data is data that needs to be, well, structured. That means there needs to be some standards about what various data elements are called, what are valid values for those elements, which way up the full stops go, and so on.

The good news is that this has been done. As a wise man once said, "The good thing about standards is that there are so many of them."

At one time, there were many competing standards for web structured data. In order to bring some order to this chaotic situation, in 2011, Bing, Google, and Yahoo (the big search engines of the time) got together to decide which standard they should stand behind. From this, the

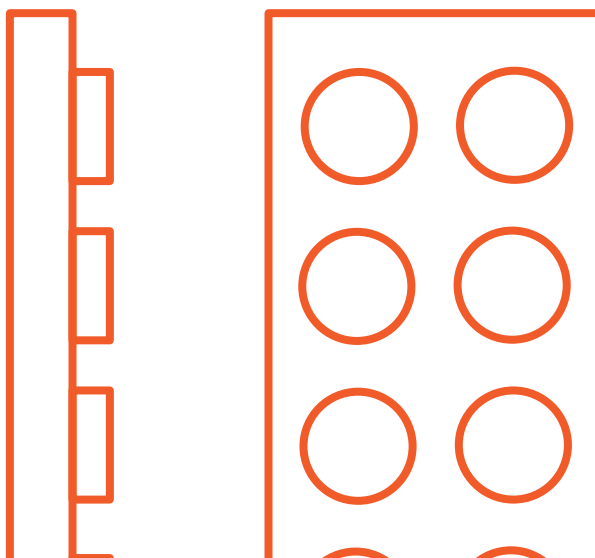
organization Schema.org was formed to be the body which would maintain and control the structured data standards.

During these discussions, three competing formats were proposed: JSON-LD, Microdata, and RDFa. The relative pros and cons of each proposal were discussed and eventually, like all good standards bodies, they adopted all three.

This is an improvement, inasmuch as now you only have a choice of three arcane, esoteric, and seemingly obscure languages to master. For the casual user, it is still confusing. So what to choose?

Like many things in life, it depends. If we assume Google will be our search engine of choice, the good news is that it supports all three standards, but prefers JSON-LD. This would then seem to be a good place to start.

2

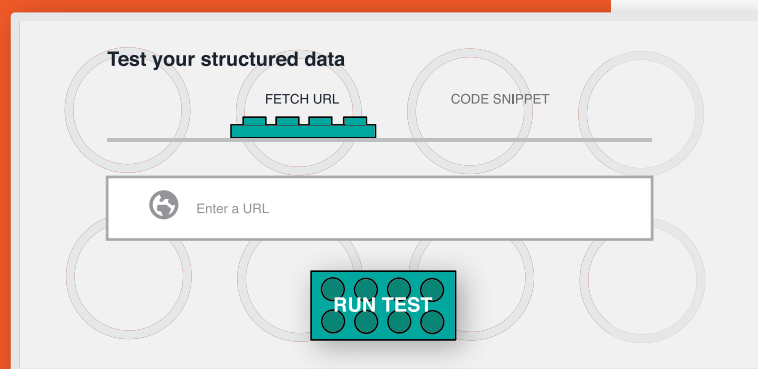


CHECKING FOR STRUCTURED DATA USAGE

Google offers a tool to check a website's use of structured data. The tool is easy to use: simply paste a URL into the search box. You will see, returned on the left-hand side, the raw HTML code, and on the right, the decoded structured data. Try the test tool on your site — you may be surprised to find some structured data there already.

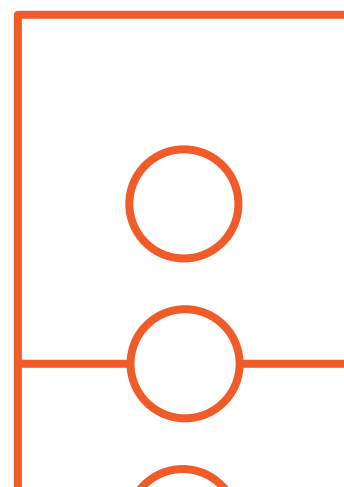
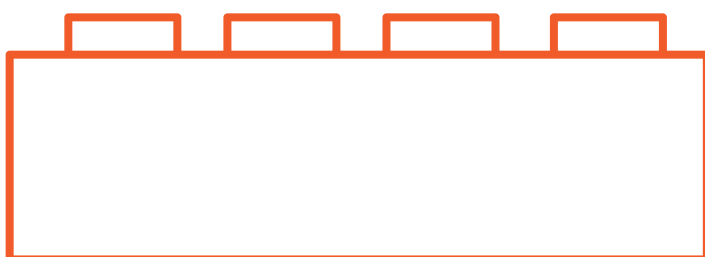
Now, I can see the structured data on my site (and others), and I can see the code that creates it on the website.

If you have had a chance to look at some examples, I can hear you saying, "I am not typing that," or some similar (perhaps rude) words to that effect. The good news is, you don't have to.



Follow the link below to access Google's structured data testing tool: pagesseomag.com/testing_tool

3



CHOOSING A STRUCTURED DATA DEV TOOL

According to data from CodeinWP.com, it would seem that WordPress powers 28.9% of the internet. True or otherwise, we will start there. A casual search of Google with a query like [WordPress structured data plugin] will turn up a great many results. Some good, some not as good, but all aimed at allowing non-technical users to create structured data that is correctly formatted.

When choosing a tool, we need a good compromise between complexity and usability. Often, a flexible, fully-featured tool will be harder to use than a less-complex but less-capable one. One such tool, “WP SEO Structured Data Schema” by Kansas City SEO, seems to find this balance.

4

PUTTING IT ALL TOGETHER

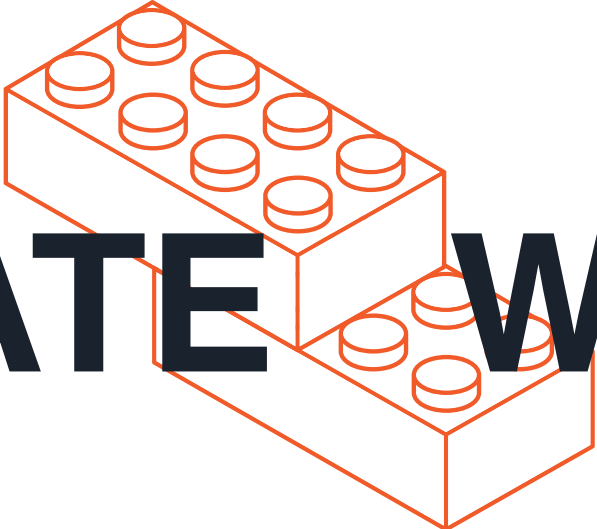


Structured data is a way to make the meaning of your content more apparent to search engines, so they can more quickly and accurately identify the content that's being presented on your website.

Structured data must be coded into your website according to the standards defined by Schema.org, but before you go putting your coding pants on, have a look around: you can easily find a tool or plugin that can help you create structured data as a part of your standard content creation workflow.

If you're going to make changes to your website by adding structured data, make sure to check your analytics before you start to get a baseline. Monitor and track any shifts in performance to ensure your changes are actually having the desired impact on traffic through the site. ©

CREATE WELL



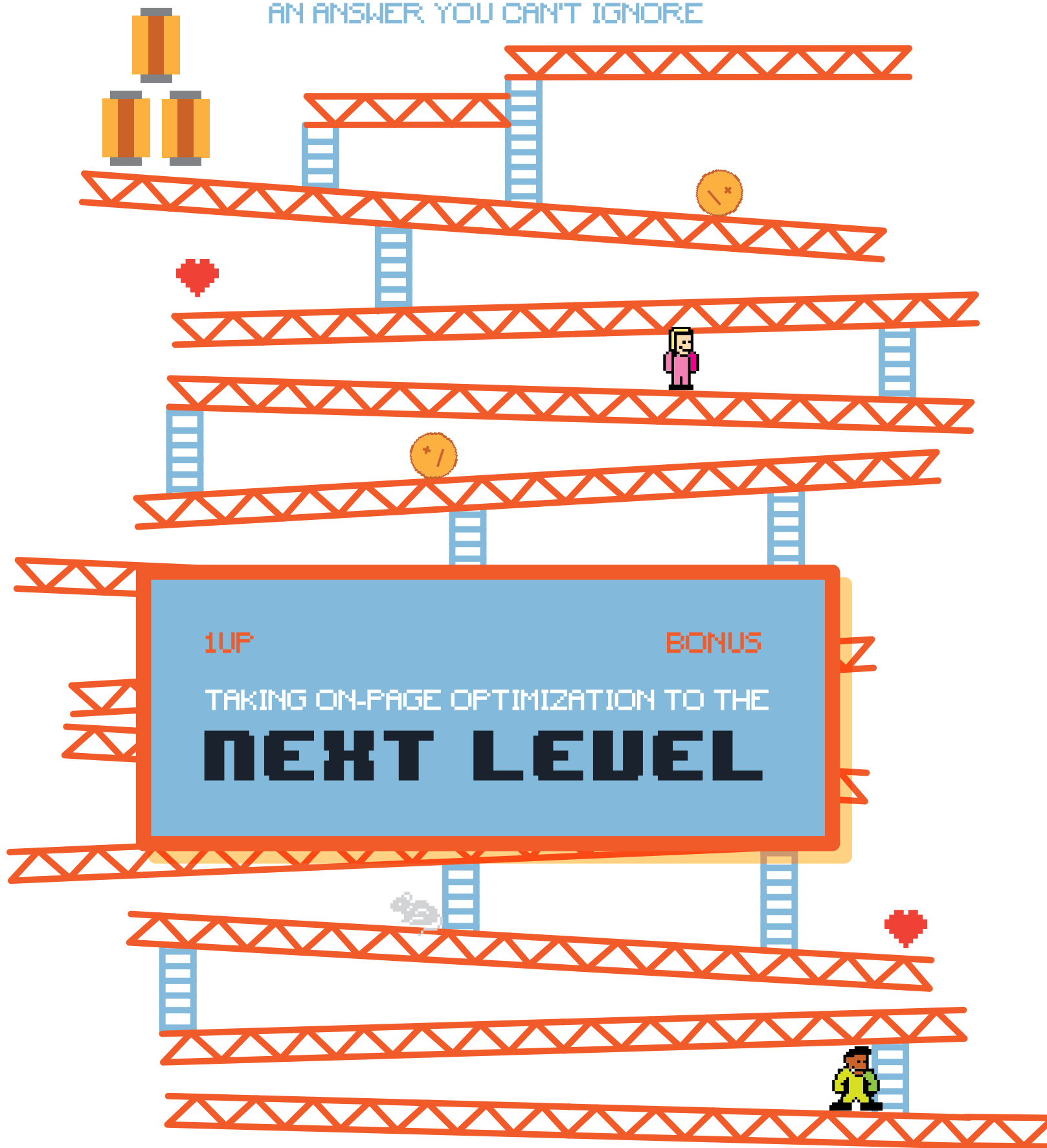
HOW DOES THIN CONTENT USUALLY HAPPEN? — MORE OFTEN THAN NOT IT IS BECAUSE A PUBLISHER KNOWS ENOUGH ABOUT SEO TO HAVE PAGES FOR DIFFERENT TOPICS TO ATTRACT MORE TRAFFIC, BUT THEY DON'T PUT IN ENOUGH EFFORT INTO THE PAGES THEY ADD TO MAKE THEM **UNIQUE** AND **VALUABLE**.

-ERIC ENGE

Stone Temple
Consulting



AN ANSWER YOU CAN'T IGNORE



CHARLIE WILLIAMS

SEO STRATEGIST, SCREAMING FROG

For all the constant advances in SEO, on-page optimization is often unsophisticated. You've optimized the Title tag, Meta description, and H1, and even some ALT tags.

Time to hit the bar?

Optimizing your page means more than the placement of your keyword targets, important as they are. It's the whole experience: showing how relevant your site is to the searcher's intent — the goal they want to achieve through a search engine.

GET CLEVER

In the last issue of *PAGES*, we learned about keyword research: finding ways our audience searches for our product, service or information. Now let's put it into practice.

One of the smartest ways to achieve better on-page optimization is increasing the number of keywords your page targets. In our post-**Panda*** world, we thankfully no longer need a separate page for each minor keyword variant.

Similar search intent is often split between many different queries, but we want our pages to be relevant for all of them. Some of these variants get called **LSI (latent semantic indexing)*** keywords, but it's enough to think of them as synonyms and closely related searches. We can find them in several different ways.

Related searches shown in search engine result pages (SERPs) are a great place to start. Next, take a look in Google Search Console to see all the queries one of your pages ranks for — often there are many you've never considered. I also recommend using keyword tools that find semantically similar terms to target. You can also use these tools to help you manually group the search queries that bring up the same pages time and again.

Several popular SEO tools have features that can help you do this. Ahrefs's Keyword Explorer identifies a 'parent' for your keywords, which groups all the variants that the same pages tend to rank for. Moz and SEMrush's tools can semantically group keyword suggestions.

If you don't have access to these tools, you can achieve a similar result by taking your research and assigning your own intent topic to each keyword. This will help you segment them into buckets.

This lets us optimize our page to increase the number of queries it can rank for. Are there terms from Search Console that we already appear for and can build on? What other keywords do competitor pages also rank for? What related topics might be of interest to our reader?

Put simply, what other searches should our page rank for, and where else could it appear if it were more comprehensive?

Increasing your organic footprint not only helps you rank for more, it's a great way to add depth and prove you should rank for your main targets.

How do we use these synonyms?

HOW TO USE KEYWORD VARIANTS

- Add major variants to our tags, or even target a more realistic variant.
- Encourage your writers to use different terms, arming them with the synonyms from your keyword and audience research.
- Answer the subtle differences for each variant through the page (helping expand our content).
- Test our Title tag and Meta description with different variants to improve click-through rate.
- Use variants in our subheadings and copy. Use the different language to structure your page.



*Google Panda -

Released: February 23, 2011

Function: Rewards high-quality websites in search results. Aims to weed out thin, low-quality, and duplicate content, while also addressing issues like high on-page ad-to-content ratios.



*LSI Keywords -

Keywords semantically related to your primary keyword. Patterns in user searches determine how one keyword is usually associated with another.





RELEVANCY ISN'T AN SEO TRICK— IT'S CREATING BETTER EXPERIENCES

ADDRESS INTENTS

The SEO cliché is “build great content,” but showcasing relevancy isn't an SEO trick — it's creating better experiences.

While our metadata is a clue to what the page is about, the content is the steak to that sizzle. Looking at the pages that are ranking well for our target search terms helps us see the nuances of the topic that search engines find important.

Analysis techniques such as **TD-IDF** (term frequency-inverse document frequency)* can tell us how important a term is within a selection of documents, for example, the top 20 ranking pages for a search query. This analysis helps us spot both the common terms that every ranking page includes (so we won't be as relevant without), and those nuances that individual pages include to show greater depth (ranking for an extra set of topic variants others miss out on).

Similarly, we can use this analysis to see co-occurrence: those terms that regularly appear with our main target term. By not addressing these, our content might seem to be lacking. Ensuring our content covers all the relevant major bases, combined with our synonym keyword research, can give us a depth that outshines the competition.

On-page SEO is about understanding intent, and building the best possible answer to the question the searcher is asking, from how to cook a lasagne to where to buy killer heels.

Google has given us clues about what constitutes a worthy answer.

**TF-IDF - Statistic that tell us how important a word is to a document (or selection of documents), typically used as a weighting factor in searches of information retrieval.*



DEMONSTRATE EXPERTISE, AUTHORITY, AND TRUSTWORTHINESS

There's been plenty of analysis of Google's Search Quality Rater's Guidelines since their release in 2015, but one of the most practical applications is their notes on Expertise, Authority, and Trustworthiness (EAT).

Taking these principles and seeing if our content stacks up tells us if we need to do more. One powerful way to get a neutral perspective is to use the Rater Guidelines examples of “High Quality Pages” as the basis for user testing questions. Does your content pass the test?

To achieve this level of quality, you need to learn to write well or hire someone who can speak to your audience. You need to keep them on your site through content that demonstrates EAT, whether through a compelling insight or superb product presentation. It can be helpful to use those audience personas other marketers put such stock in, or speak to those (the sales team, customer services) who know your audience the best. You must also consider which content formats are being rewarded in your niche.

Good structure and depth might not always directly influence visibility, but better UX will get you more clicks, longer dwell time, and even more links — all of which help SEO, and conversions (and who doesn't like those?).

And of course, don't forget to include links. No page can answer every aspect of a query, so including links to other pages, both internal (highlighting our breadth of knowledge, as well your expertise) and external, can show your page is a great answer in itself and as a hub to further reading.

This is how we build that elusive “10x” content. No tricks, just a better answer, based on topic research.

ASK THE AUDIENCE

Want to add some depth to your pages without having to write a word? Let your audience help.

User-generated content (UGC) is a way to quickly add not just more content to your page, but also depth and expertise that you might not possess. Google has said that “a healthy, thriving community” can signal your site as an authority. As we’ve seen content like comments, questions, and reviews ranking, UGC can also help us be relevant for even more long-tail queries.

For e-commerce pages, this can be very powerful. When it is hard to differentiate between suppliers, an engaged audience can showcase your authority and trustworthiness, making your pages more useful than the competition.

Not only does it add depth to content, we can analyze UGC for patterns — common terminology we can optimize for, and the questions our main content should answer. If users are asking the question on your site, it’s likely others are typing it into Google.

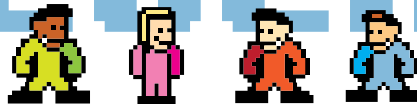
CLOSING THOUGHTS

Search engines interpret a query and decide what the best answers could be. Google doesn’t know the answers to everything (yet...), but if we don’t match their interpretation of intent, no amount of keyword optimization alone will work.

That is the challenge of building a page for SEO today.

Your job is to balance on-page optimization (more honestly, content optimization) with user experience, proving you are a vital resource users will love and search engines can’t ignore. 🎯

LEVEL UP UNLOCKED



PRESS START

PEOPLE SOMETIMES JUST OPTIMIZE THEIR WEBSITE ON THE SURFACE, BUT FAIL TO **DIG DEEPER** AND MINIFY CODE, OPTIMIZE THE SERVER, RELOCATE JAVASCRIPT TO THE FOOTER, REMOVE UNNECESSARY TRACKING SCRIPTS NOBODY IS USING AND SO ON.

These tweaks, when accumulated, will produce a markedly better user experience in terms of page load speed, and this is a direct signal for improved search engine rankings.

-**RODOMIR BASTA**
Four Dots





RACE

TO THE TOP



QUICK TIPS FOR A FASTER PAGE

WES MARSH

DIRECTOR OF DIGITAL MARKETING, SOLODEV

“Slow and steady” always wins the proverbial race, but Aesop might rewrite his fable if he saw how fast internet speeds are shaping online success.

For the fourth consecutive year, mobile devices remain the go-to platform for users to visit their websites of choice. The average time for mobile websites to fully load is 22 seconds, according to Hubspot data. However, the average web visitor will start to leave after only 3. It's time to find a way to slim down everything on your website: Google's Speed Update is coming.

Google announced in January that page speed would weigh more heavily into mobile search ranking. As the Google Webmaster Central Blog points out, the updates always start by affecting the slowest websites first — but don't get comfortable. Just because it affects a small percentage of queries right now doesn't give you an excuse to wait. Google has already offered a six-month heads up to prepare for this update. It's fair to assume that the Speed Update will continue to be a transformative part of the core algorithm.

HERE ARE FIVE TIPS FOR STAYING ONE STEP AHEAD OF THE SPEED UPDATE, EVEN IF YOUR WEBSITE FEELS MORE LIKE THE TORTOISE THAN THE HARE.

USE A CDN

Content delivery networks (CDNs) often improve performance by reducing latency while also offering scalability for peak traffic times. In short, they'll speed up your site. CDNs work by distributing your website files across a network of servers around the world. Just like with online shipping, closer proximity to your desired product means it shows up on your doorstep faster.



REDUCE AS MUCH AS YOU CAN

A visually-appealing site is one that is often, unfortunately, slow to load. Image size, Javascript, CSS, plugins, animations, and anything else that can be culled out of the page, while still delivering a quality UX, should be made as small as necessary. Speed is the goal, but don't over-optimize for speed and leave your users with a poor experience.

LAZY-LOAD YOUR PAGES

Page speed is all about the time it takes to first render. We're already seeing many websites load only the header and the first 100-150 words of content, then prompting users to click to "Read More," or load more of the page's content. This helps improve speeds by not having to complete the entire rendering of the site all at once. However, make sure you put links into your content so search engine crawlers continue on to index the lazy-loaded content.

MANAGE YOUR TAGS

These little bits of code can be put to use for a variety of tracking and analytics purposes, but they can add a lot of bulk to your site. Google offers a free solution to manage all of your tags. In one script, you can hold them all — from Universal Analytics, AdWords, Facebook Pixel, and other tags for remarketing, call tracking, A/B testing, and events. More importantly, by selecting the specific triggers required to fire those tags, you improve the speed of your site by not having individual scripts load on each page.

USE THE LIGHTHOUSE CHROME EXTENSION

This is another free tool from Google, and it audits websites based on their optimization for Google ranking factors, making it one of the most effective site review tools available. In a matter of seconds, you can get a snapshot report of your site performance, accessibility, SEO checks, and adherence to best practices. This should be a staple in your toolbox along with GTMetrix, PageSpeed Insights, and your other favorite speed tools.



WINNERS CIRCLE

Even if your page seems speedy, try out each of these tips for a faster site. Though providing a great experience for your visitors should be incentive enough, page speed isn't simply a UX factor now that Google is shifting to meet user demand. The extra work you put into greasing the wheels could soon be the difference between you and your competitor at the top of the SERPs — don't get left in the dust. ©

An aerial view of a city skyline at sunset, with numerous skyscrapers and buildings. The sky is a mix of orange, yellow, and blue. The text is overlaid on the image.

**IF YOU TRACK TOO FEW
RANKINGS IT WILL GIVE
YOU A TINY FRACTION
OF THE STORY – TRACK
TOO MANY AND IT'S
EASY TO SEE GOOD
NEWS OR BAD NEWS
EVERYWHERE YOU LOOK,
WITHOUT EVER REALLY
MAKING A DIFFERENCE
TO THE BOTTOM LINE.**

-STEPHEN KENWRIGHT

*Strategy Director,
Branded3*



WHEN TWO THINGS SEEM THE SAME, HOW CAN YOU TELL WHICH IS BETTER?

This is the conundrum that search engines face in dealing with duplicate content — pieces of “appreciably similar” content on the web.

As you can probably imagine, it’s certainly not uncommon to find similar pieces of content on the web, or even on your own website. There are plenty of instances when web users’ natural, normal behaviors on the web lead to multiple pieces of content that look the same, and most of the time, it’s not intentional.

For example, think about boilerplate content — the kind of “standard” content that exists on nearly all the pages on your site, including elements like sitewide navigation bars, headers, footers. All of this is technically duplicate content, but it won’t hurt your optimization. However, there are some types of duplicate content that cause search engines a lot of confusion and impact the optimization of a site.

Duplicate content is one of the most common issues on websites. According to a study conducted by Raven Tools, 29% of all pages on the web have duplicate content. Problems with duplicate content can be complex to navigate, but it’s not too tough to figure out if you need to take action. If you do find your site is being impacted, there are steps you can take to fix it.

HOW IT HAPPENS

Usually, major duplicate content issues arise because of something technical on your site that’s gone wrong.

One of the most common ways that duplicate content happens is when there are multiple versions of one site. For instance, if a site has www and non-www versions, or both http:// and https:// versions. Printer-friendly versions of pages, inconsistent URLs and URL parameters, and session IDs can also lead to duplicate pages.

Another common cause is copied content. E-commerce retail sites are particularly prone to this issue, as it’s common for retailers to use the manufacturer’s descriptions of products on their sites. If other retailers also use that same description, the content becomes duplicated across the web.

WHAT IT DOES

Search engines must determine which version of a piece of duplicate content to include in search results. After all, there’s little benefit to users to encounter a SERP full of seemingly-identical pieces of content.

To improve UX, Google has built-in ways of handling duplicate content, and it manifests like this:

“In order to show you the most relevant results, we have omitted some entries very similar to the [X number] already displayed. If you like, you can repeat the search with the omitted results included.”

Duplicate content causes a number of issues that can impact the performance of your site.

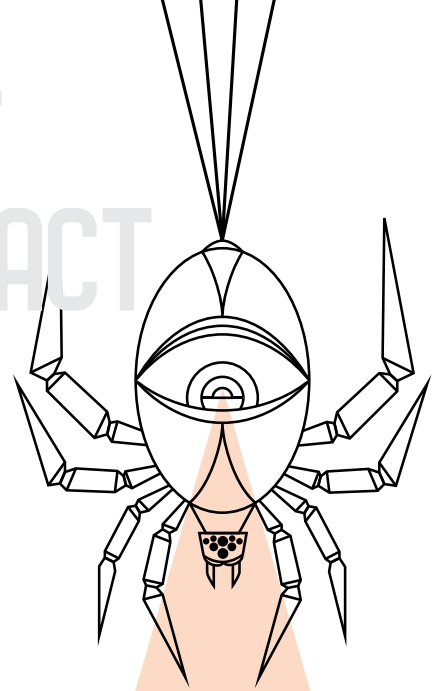
While duplicate content isn’t penalized by Google’s algorithm, it still impacts your rankings: duplicate content eats up your crawl budget, as spiders must do extra work on your site sorting the content and deciding which pages to rank.

If you’ve intentionally duplicated content in a malicious way (for example, through plagiarism or to intentionally manipulate search results), you can be subject to a manual penalty — that’s the kind a person puts on your site. This is also another item you can add to your list of reasons to avoid using spun content on your site. To have this kind of penalty removed, you’ll need to begin an appeal process that can be conducted through Google Search Console.

Additionally, because duplicate content forces a search engine to choose which page to show users and which to hide, the version included in the SERPs isn’t always optimal.

Duplicate content also can cause link equity to become diluted if sites linking back to your page use different versions of the URL — for example, yoursite.com instead of www.yoursite.com. This means the authority you gain from incoming links is distributed throughout the different versions of your site, instead of being directed to the version of the page you want visitors landing on. Having multiple versions of the pages on your site can also impact tracking of engagement signals, such as bounce rate, time on page, and page views, further negatively influencing ranking factors.

DUPLICATE CONTENT
CAUSES A NUMBER OF
ISSUES THAT CAN IMPACT
THE PERFORMANCE
OF YOUR SITE.



HOW TO DEAL

Understanding how to handle duplicate content on your site can improve optimization and rankings. Luckily, since it's a common issue, managing duplicate content is rather straightforward. Google has created a variety of solutions to help webmasters discover duplicate content issues.

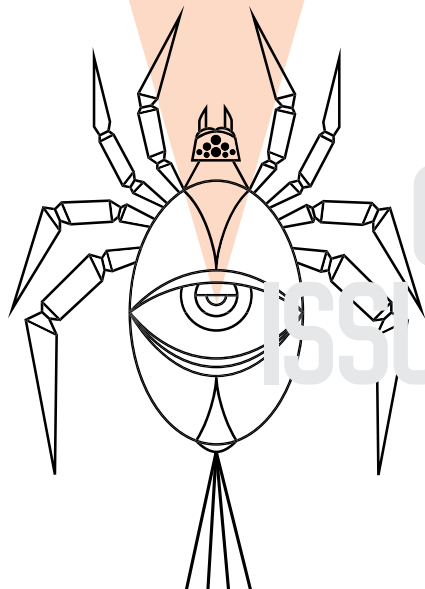
Duplicate content is best handled by making sure it isn't generated in the first place.

These steps can help you avoid creating duplicate content unintentionally:

- Disable session IDs
- Get rid of printer-friendly pages
- Build a script to keep URL parameters consistent
- Avoid using URL parameters for tracking links
- Disable comment pagination in WordPress
- Create unique content for international versions of your site in advance — don't clone those pages, even if each version of the site is in the same language

But don't fret — if you've already got issues with duplicate content, or suspect you may, there are solutions available.

The first step is to head into Google Search Console and check out what it has to say. GSC will give you a rundown of any of your URLs with duplicate titles or meta descriptions so you can start making fixes. However, this doesn't always give you the full picture, particularly if you have pieces of duplicate content on your site that happen to have different titles — this usually happens with content that's listed under multiple categories or tags on your site, like blog posts.



The next step is to do some good old-fashioned searches, putting some helpful search operators to work. The more specific you can be with your keywords, the better results you'll get. If you want to find all of the URLs on your site that include the same piece of content, try this:

Site:[yoursite.com] intitle:[your keyword]

You can expand your search across the web as well, using:

intitle: [your keyword]

Canonical issues — the kind where there's multiple URLs headed to the same spot — can be fixed by repairing any 301 redirects that are improperly placed. You can also utilize a canonical tag to repair this kind of issue, which tells search engines which version of your site is the “real” copy and where to assign the link equity. Another tag that can be helpful is the *Meta noindex, follow* tag, which allows a search engine to crawl the links on a page, but not to index them, which is helpful for avoiding issues that could be caused by restricting crawl access to parts of your site.

Remember: not every instance of duplicate content can be “fixed,” but repairing the ones that can will improve your optimization. Once you've taken care of any outstanding problems, set up safeguards to prevent future issues and regularly audit your site to check for new issues. Dealing with duplicate content will set the foundation for getting the full potential out of your site — don't let it hold you back. 🕸

DUPLICATE CONTENT
CAUSES A NUMBER OF
ISSUES THAT CAN IMPACT
THE PERFORMANCE
OF YOUR SITE.

LÄNKARATI

Ready-To-Assemble SEO



Masterminds
x 4



Contributors
x 50



Coffee a Day
x 4

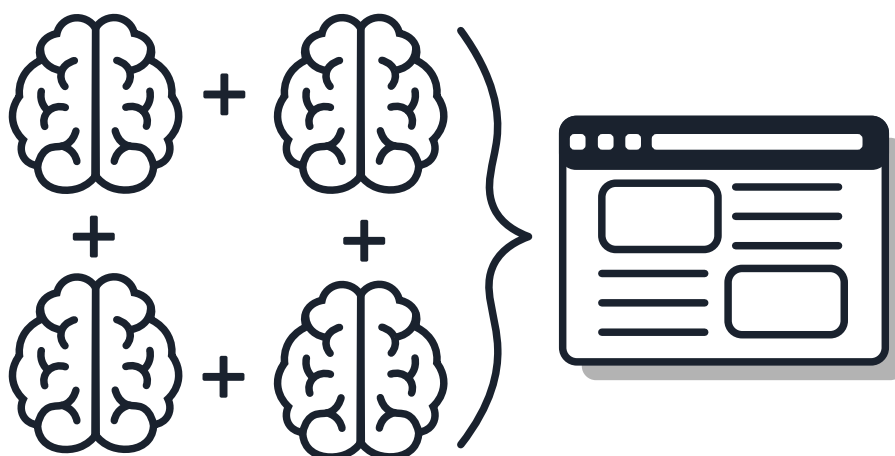


Website
x 1

1



2



3



4



CRAWLERS AND SEARCH ENGINES WILL DO THEIR BEST TO CRAWL, EXECUTE, AND INTERPRET YOUR JAVASCRIPT, BUT IT IS NOT GUARANTEED.

Make sure your content is crawlable, obtainable, and isn't developing site latency obstructions. The key: every situation demands testing. Based on the results, evaluate potential solutions.

-ALEXIS SANDERS
MERKLE | IMPAQT







SOLVING CRAWLER OBSTACLES

**ALL DECISIONS MADE IN SEO ARE
BASED ON INFORMATION THAT
REQUIRES CAREFUL COLLECTION.
GUIDANCE MUST COME FROM
EDUCATED GUESSES, INFORMED
BY TRUSTWORTHY DATA.**

One of the best tools for retrieving this data is **Screaming Frog**, a crawler which can be tweaked in many different ways to accomplish specific tasks of data gathering. However, ensuring the integrity of your data is rarely a straightforward path.

There are many potential obstacles in the journey towards information enlightenment — here, we'll discuss some of the more common issues you might run into.

- NICHOLAS CHIMONAS -
Director of SEO, WTP Inc.



THE SITE IS TOO BIG

Screaming Frog runs off of your system memory and is thus limited by the capacity of your machine.

Generally speaking, I run into issues with a plain text crawl after around ~75k-~100k URLs using a machine with 16 GB of RAM and a 2.2 i7 processor.

It's possible to run Screaming Frog on a server remotely to increase the power available. This is most easily accomplished through Amazon Web Servers EC2 instances, which scale in processor and RAM to your desired definition.

For a full walkthrough, check out Mike King's post from ipullrank.com in the resource box below.

If you are more comfortable using Google Cloud Compute engine, I can verify Screaming Frog works there too. However, my experience with this was a whole Linux ordeal and, while it was interesting and informative, I'll never do it again. Amazon Web Servers is a much simpler solution.

Another possibly quicker solution, if you only occasionally deal with large domains, is to crawl in segments using Screaming Frog's include/exclude features, disallow "crawl outside of start folder" in settings, and engage in several crawls of all main subfolders, presuming each subfolder is within your RAM's capacity.

As a side note, you can increase Screaming Frog's available RAM allocation (the standard is 1 GB for a 32-bit machine, and 2 GB for 64-bit). If you use all your allocated memory, you'll receive a warning to increase your memory or it will become unstable.

TOO BIG!



How To Run Screaming Frog And URL Profiler On Amazon Web Services by Mike King.



pagesSEOMag.com/how2ipullrank



THE CRAWLER IS BLOCKED BY THE SERVER

There are many evil crawlers on the internet, full of malice and server load. Large domains tend to take an inexorable approach, allowing entry only to those defined as pure by the royal domain. All others shall receive the forbidden gates of 403.

There are many potential causes for this, and it behooves you to determine exactly which cause is to blame before approaching the domain's gatekeeper with a request for entry — specificity is appreciated by technical individuals.

POSSIBLE CAUSES TO TEST:

User-Agent Whitelist or Blacklist

Try crawling as Chrome, but always use a VPN first to ensure you haven't already been blacklisted by IP after approaching the server identified as a crawler. I always prefer to identify as a crawler in my first crawl in case user-agent black listing is inactive, because then the server logs will be cleaner later. It's not a huge pain to filter your IP out later, but I'm constantly on VPNs and it can be a mild pain.

Server Rate Limiting

Crawling too quickly can upset a server, as bots can be used for distributed denial of service (DDOS) to crash the system. You may receive status code 403 errors (access forbidden) after a decent number of URLs initially responded 200 OK — this is usually because of rate limiting.

Try slowing your crawl to under 5 URLs per second and limit threads to 1 or 2: this seems to be the most common acceptable crawl rate I've found in my work. This method takes much longer for huge domains; do the math and run it overnight, the weekend, or on a server.

For example, a 1 million page site at an average of 5 URLs per second will take a couple of days to crawl:

86,400 seconds per day x 5 URLs per second = 432,000 URLs crawled per day.

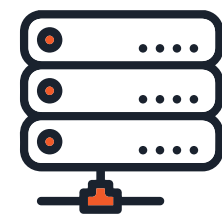
IP Whitelist or Blacklist

If you've already tried changing user-agent, limiting your crawl rate, and changing your IP via VPN, you're probably dealing with a server that has a whitelist. It's also possible that you've already landed yourself on the site's blacklist by being too liberal with your crawl testing.

At this point, you can approach the gatekeeper with the steps you've attempted, and they will most likely happily put your IP on a whitelist, because you know how to crawl respectfully.

Unless, of course, you did get yourself blacklisted. Then you may be heckled.

B L O C K E D



THE CRAWLER CAN'T SEE THE DESIRED CONTENT.

Sometimes the HTML source is different than the rendered document object model (DOM). Observe the differences between “view page source” and “inspect element” (which reveals the rendered DOM). Usually, this is because of fancy new JavaScript libraries which modify the DOM, like AngularJS, ReactJS, and jQuery.

Fortunately, Screaming Frog can utilize the Chromium Project library as a rendering engine — modify Screaming Frog’s configurations in the “rendering” tab to render JavaScript. It is slower to crawl this way, so only do it when necessary, but it’s becoming more necessary with more domains every day.

However, Screaming Frog cannot do it all, and its JavaScript rendering solution is not perfect.

Certain data will not appear in the rendered DOM unless the “check external links” setting is enabled when specific elements you need to extract from the page are delivered through JavaScript. When combined with major domains that implement “welcome pages” and/or custom tag tracking URLs (like Forbes), it’s impossible to ignore JavaScript-based redirects to content while also allowing JavaScript-rendered content to appear in the extractable HTML during crawl.

This is a conundrum when specific data, like a view count per post, needs to be extracted from a large number of URLs from an external domain; for example, perhaps one in which you have a guest column or competitive analysis interest in.

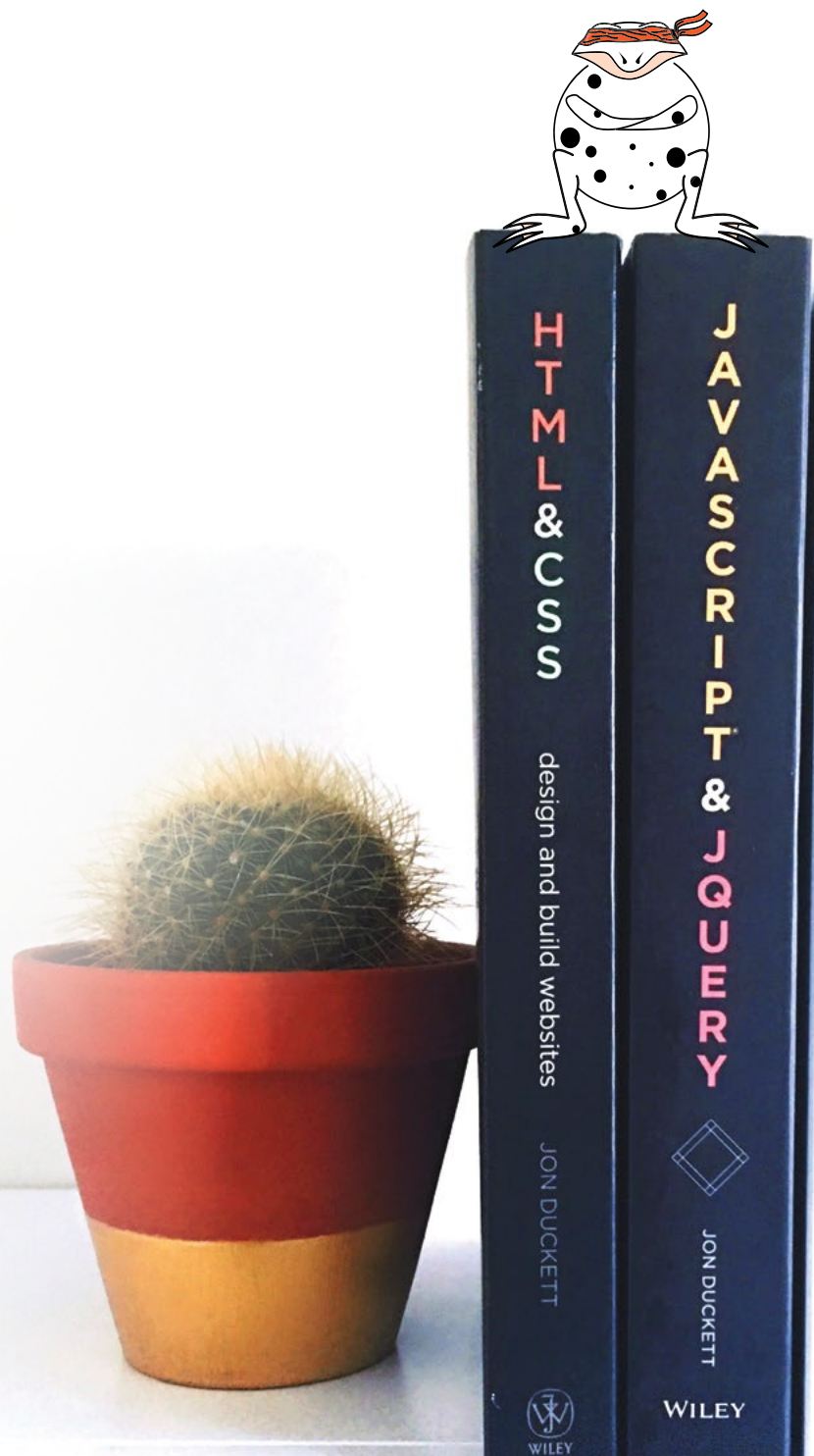
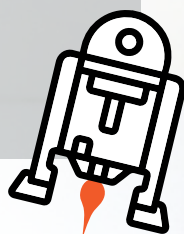
Enter **artoo.js**, a client-side scraping companion. This is the cleanest solution I have found to extracting rendered DOM quickly and at scale. Artoo automatically injects jQuery and exports in pretty JSON.

GG, WP. Happy crawlings. 🕸

Follow the link or scan the QR code to access **artoo.js**.



pagesSEOMag.com/artoo





Your organic search solution. Learn more at pageonepower.com



PAGEONEPOWER

THE INDEX LANDSCAPE

THIS YEAR HAS PROVEN TO BE A COMPETITIVE ONE IN THE SEO SOFTWARE TOOLS INDUSTRY.

Several of the major players in search analytics software have made big changes that impact their backlink indexes over the last year.

Backlinks are one of the most important ranking factors for a site. As such, tools that provide accurate data about the backlink profiles of sites are highly valuable to SEOs — after all, this data is the springboard for most of an SEO's work.

Backlink indexes aim to map the links on the web in a way that's parallel to Google's own search index, providing webmasters information about metrics like traffic, keywords, and competitor performance. These indexes are powered by crawlers, designed to operate as similarly as possible to Google's own Googlebot, which regularly explore the web and gather data about sites.

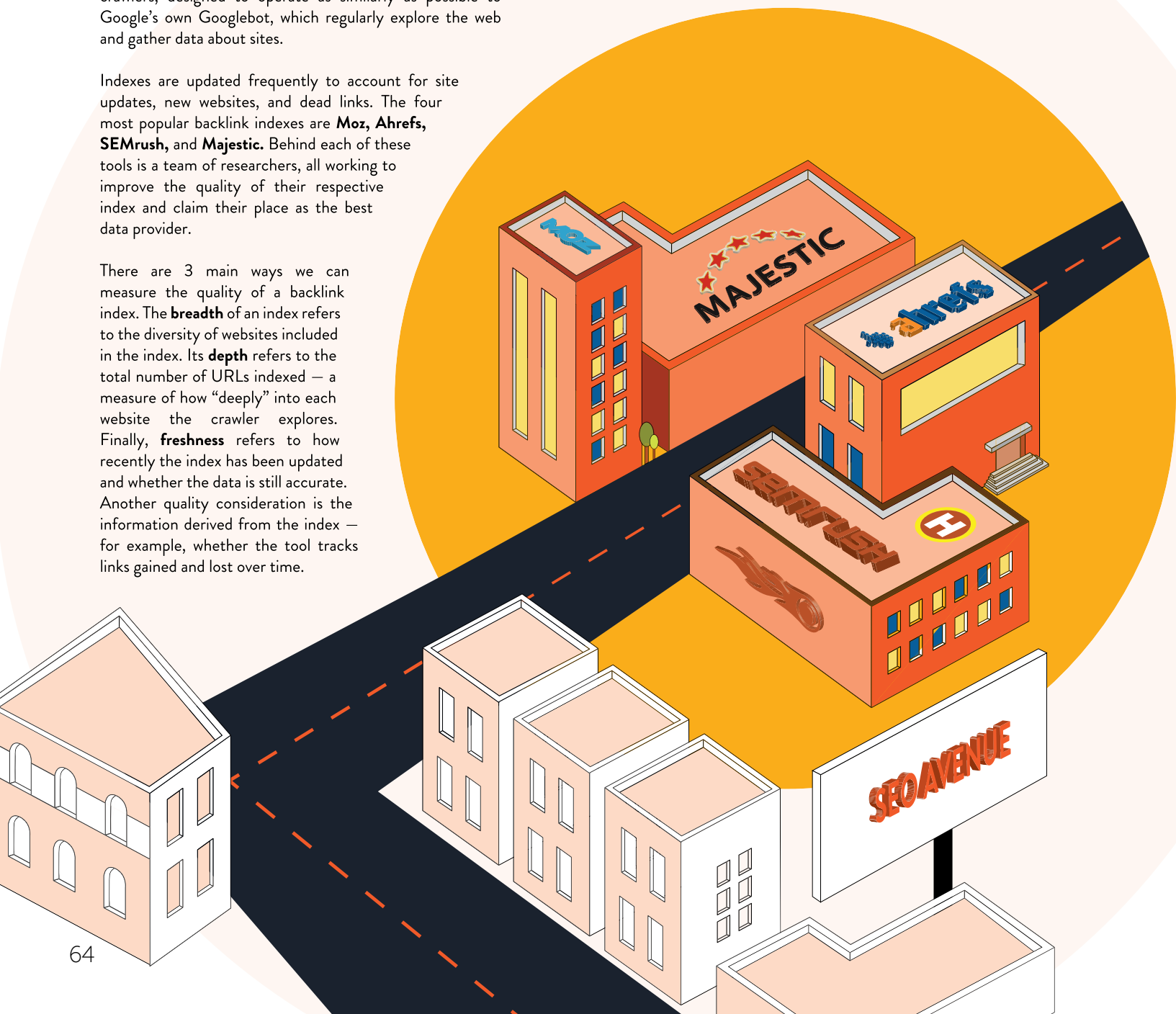
Indexes are updated frequently to account for site updates, new websites, and dead links. The four most popular backlink indexes are **Moz**, **Ahrefs**, **SEMrush**, and **Majestic**. Behind each of these tools is a team of researchers, all working to improve the quality of their respective index and claim their place as the best data provider.

There are 3 main ways we can measure the quality of a backlink index. The **breadth** of an index refers to the diversity of websites included in the index. Its **depth** refers to the total number of URLs indexed — a measure of how “deeply” into each website the crawler explores. Finally, **freshness** refers to how recently the index has been updated and whether the data is still accurate. Another quality consideration is the information derived from the index — for example, whether the tool tracks links gained and lost over time.

We can also consider the function of the crawlers that power each index, attributes such as **crawler politeness**: the likelihood that a crawler will be blocked from indexing a given site. When a crawler visits a site, it requires resources (bandwidth) from the hosting site. While most webmasters are permissive of crawler activity, overly-aggressive crawlers can get banned from websites for not utilizing these resources judiciously, meaning backlink data software providers have an incentive to make their crawlers as polite as possible to avoid skewing the index because of crawler bans.

All of these metrics are important to consider when you're basing business decisions on backlink index data.

This year's competitive market environment means we've seen improvements made by several major providers as they've been driven to utilize better research methods to improve the quality of their indices. If you're looking for SEO software and keyword research tools that utilize a backlink index, know that there are plenty of options available, but that you'll need to weigh them carefully to decide which one is best for your budget and needs. 🎯



TECHNICAL GLOSSARY

A FEW TERMS YOU MIGHT WANT TO LEARN

Accelerated Mobile Pages (AMP) - an open-source initiative aimed at making webpages load near instantly, improving mobile user experience

Crawler (spider) - an internet bot that systematically browses the web, typically for the purpose of indexing. Can be utilized for a variety of data-gathering purposes, e.g. for scraping — extracting and copying web data.

Crawl budget - the number of pages on your website that Google crawls on a given day

rel=canonical - an HTML element that indicates that a version of a webpage is the “preferred” one, i.e. the original version

JavaScript - a programming language that allows for the creation of interactive webpages

Log files - files that are a record of events, messages, or processes occurring during the execution of a system

Pagination - breaking a document into separate pages. Comment pagination — creating a new webpage for each comment on a blog post or other piece of web content — is one of the most common ways that duplicate content gets created on the web.

Schema markup - code you can add to your HTML which helps search engines provide users more informative results about your page in search engine results pages, for example, information like product prices, the date an article was published, or review information

Site architecture - the planning and design of the functional components of a website. SEO considerations include UX, search engine accessibility, sitemaps, and the way a website is structured with internal links (typically determined according to relevancy: relevant pages will be linked together).

Sitemap - a list of the pages of a website. Sitemaps can either be human-readable to help users navigate a website, or XML Sitemaps, which are structured listings of the pages in a site that inform search engines about the important content on the site.

Top Level Domain (TLD) - the last segment of a domain name that follows immediately after the “dot” symbol, e.g.: .com, .net, .edu, .au

Universal Resource Identifier (URI) - a sequence of characters that identifies an abstract or physical resource. The most common type of URIs are universal resource locators (URLs), but universal resource names (URNs) are another type of URI — for example, ISBN numbers used to identify books are a type of URN.

Universal Resource Locator (URL) - the type of URI that identifies the location of a webpage (it's web address)

Hypertext Markup Language (HTML) - the standard markup language for creating webpages:

- **Hypertext** - the method by which you move around the internet, clicking on links
- **Markup** - is what HTML tags do to the text inside them: they mark the text as a certain type of text (bold or italic text, for example)
- **Language** - HTML is a language that has code words and syntax, like other languages



Googlebot: the one spider you don't want to kill.



Illustration and comic courtesy of Amethyst Tagney of Page One Power.

STAY LEARNT 



LOOK HERE

- THAT'S ONE MORE PAIR OF EYES THAT COULD HAVE SEEN YOUR AD -

FOR MORE INFORMATION ON ADVERTISING WITH *PAGES*, CONTACT US AT:

INFO@pagesSEOmagazine.com

WHAT DO YOU HAVE TO SAY ABOUT SEARCH?

CONTRIBUTE

TO PAGES

**PAGES IS A
RESOURCE
FOR DIGITAL
MARKETERS
WHO WANT
TO LEARN
ABOUT **SEO.****

**GOT SOME KNOWLEDGE TO DROP ABOUT
SEARCH?**

LET'S HEAR IT.

PAGES is powered by contributions from those who are passionate about search. Each issue is focused on a different SEO topic, but always aims to explain how SEO can benefit a business. We're looking for insightful and accessible articles that explore the SEO tactics that help us achieve those benefits, aiding readers in developing a deeper understanding of how SEO fits into a marketing plan.

If you've got something you'd like to share about SEO with digital marketers, let us know!

Visit the *PAGES* website to view our contributor guidelines, and email us to pitch your ideas or get more details about upcoming themes:

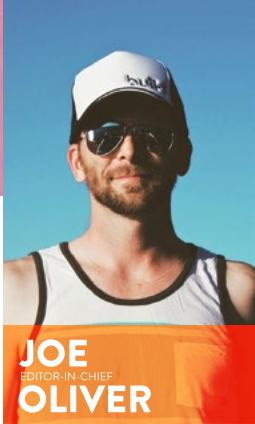
info@pagesseomagazine.com

www.pagesseomagazine.com



HI.
HELLO.
HEY, HOW ARE YA?

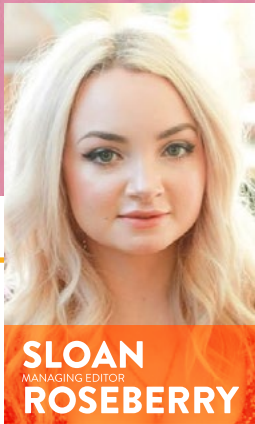
MEET THE PAGES CREW



JOE
EDITOR-IN-CHIEF
OLIVER

An Apple fanboy, camping enthusiast, and developer of leaders. Joe believes great marketing lives at the crossroads of compelling design and interesting stories. When Joe's not helping the PAGES crew live at that crossroad, you'll find him in the mountains with his family or giving his money to whatever Marvel Studios movie is currently playing.

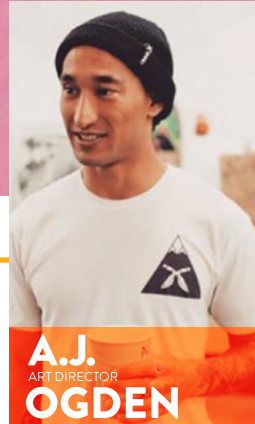
Twitter: @JoeOliver



SLOAN
MANAGING EDITOR
ROSEBERRY

Web nerd and writer who decided to make that skill marketable with a marketing degree. She's been attached to a computer since 1996, when her dad would let her watch him play Diablo on Windows 95, and loves to explore the ways we change and are changed by the web.

[linkedin.com/in/vsloanroseberry](https://www.linkedin.com/in/vsloanroseberry)



A.J.
ART DIRECTOR
OGDEN

All-weather winter enthusiast, fair-weather golfer, summer-weather kook. Prefers standing sideways on boards and playing with dogs. Enjoys trying to find Neverland and tracking coordinates with a pencil and notebook in hand. Pleasantly surprised to be pushing analog treasures in a digital industry.

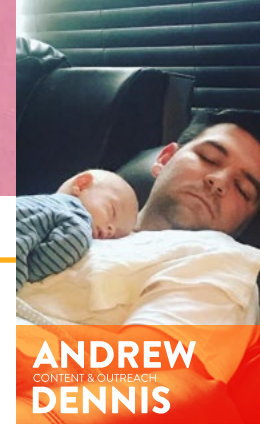
IG: @juicemachinedesign



JAN
HEAD OF ADVERTISING & PR
SCHMIDT

Wrangler of tasks, keeper of peace, and organizer extraordinaire. Social media master and knick-knack emporium curator. Knows good marketing comes from collaboration, a warm office, and a few cups of strong coffee.

IG: @tomatobaby lifestyle



ANDREW
CONTENT & OUTREACH
DENNIS

Proud graduate of the University of Idaho and a (less) proud Detroit Lions fan. Fresh member to the Rad Dad Club.

Twitter: @AndrewDennis33





THANK YOU

To This Issue's Contributors

Going Meta On Your SEO

Peter Meyers - Moz

Diagnosing Traffic Drops

Marie Haynes - Marie Haynes Consulting, Inc.

Creating Context With Structured Data

Alan Ibbett - Catholic Education Office in the Diocese of Wollongong

An Answer You Can't Ignore

Charlie Williams - Screaming Frog

Race To The Top

Wes Marsh - SoloDev

Solving Crawler Issues

Nicholas Chimonas - WTP Inc.





U F Z C N J N M H K T N E T N O C T V K
 Y F Q M C L R R F Q J M I I C C R I M N
 Q E R V H O H A X N J M V T T O B T G C
 T S I U W Y E F N Q S U K A X O P L E K
 M I H L P H P S T K B G Q C A C C E U H
 R T B R U C K E L I I B R L N C E T S W
 C R T Q B J G A R A D N A R A U S A S T
 G E J K L V R Y W L C U G R L R S G E E
 G P S M L U W T B D I I A J Y R Z L N H
 N X I E B O N I F P F N N J T A I C I Z
 I E M T I B N R U A O F K H I N T O H D
 K H M A G S P O N M F Y D Z C C D M T B
 C A Q W V S T H I E Z A H T S E K P R O
 A U E U C Z N T M T F P N H U E T E O B
 R U G R I E E U T I A E U S L R K T W N
 T P A W X F T A T S T N A Y F F P I T I
 U P P S F Q N I I A L M I A W P F T S A
 E L A C X Q I T L I P D S G G V J I U M
 H D T H I N B B O Q B A S Z A A V O R O
 E E V G D Y Y F I N I M M X D P K N T D

ANALYTICS
 COMPETITION
 DOMAIN
 INTENT
 MINIFY
 SCRAPE
 THIN

AUDIT
 CONTENT
 EXPERTISE
 LATENT
 PAGINATION
 SITEMAP
 TITLETAG

AUTHORITY
 COOCCURRENCE
 HYPERLINK
 META
 RANKING
 TRACKING
 TRUSTWORTHINESS

THANKS FOR READING! HERE'S TO YOU!

PAGES is published each quarter by Page One Power LLC,
 for a total of 4 yearly issues.

For reprints and advertising inquiries, please contact
 Jan Schmidt at jan@pagesSEOmagine.com

If you are interested in becoming a contributor to PAGES,
 you can reach our editor at sloan@pagesSEOmagine.com

Write to us at:
 Page One Power
 ATTN: PAGES Magazine
 3100 N Lakeharbor Ln, Suite 254
 Boise, ID 83703

www.pagesSEOmagine.com
 PRINTED IN THE USA
 Photography courtesy of UNSPLASH.com
 unless otherwise cited.

