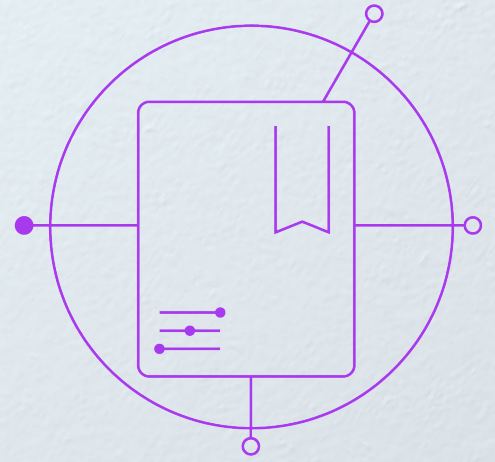


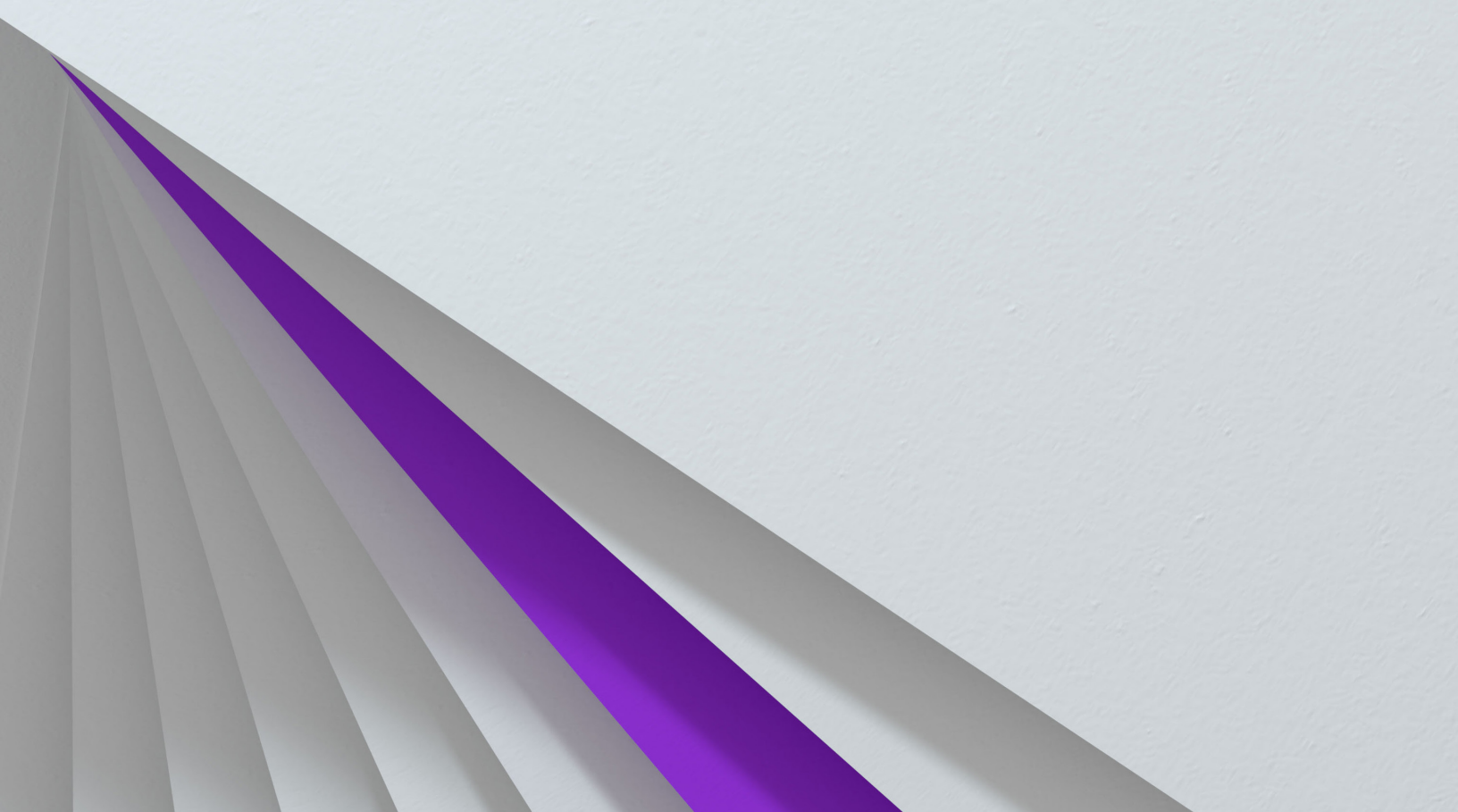
botify

PLAYBOOK



The Publisher SEO Playbook

DEVELOPED IN
COLLABORATION WITH



Dear Readers,

Digital publishers today - especially enterprise publishers - manage some of the largest, most-frequently updated websites on the internet. Regardless of publication type, getting relevant, engaging, and unique content developed and published is only half the battle. Ensuring that the content is discovered, indexed, and read can make or break its success.

Among many other things, publisher success requires editorial excellence, precise internal coordination, and of course, a reliable technology stack, core to which is a high-performing content management system, as well as a complementary, SEO platform, both of which are designed for speed and complexity.

That is why we decided to develop this publisher SEO playbook in collaboration with our partner, WordPress VIP. The WordPress brand is ubiquitous in the world of digital content management, hosting millions of websites representing nearly 40% of the web. WordPress VIP is designed specifically to serve their largest, highest-trafficked websites; focusing on enterprise and large-scale content management, offering powerful hosting, flexible infrastructure, and unmatched performance and security.

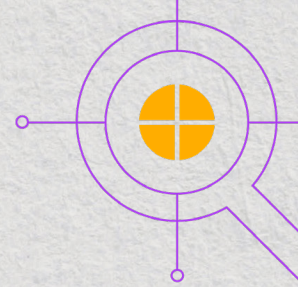
In fact, WordPress VIP powers the websites of some of the most recognizable media brands in the world such as News Corp, CNN, Penske Media Group, Hachette Book Group, TechCrunch, and more, so who better to contribute content marketing best practices and considerations, as well as first-hand examples of publisher success.

“As a Botify technology partner, we’re ready to help enterprise publishers drive growth through their content. We’re fiercely dedicated to empowering the world’s biggest companies to easily create meaningful digital experiences.”

-Ryan Sholin, Director of Business Development, WordPress VIP

Happy reading.

Your friends at Botify



What's in this Playbook

Content is King, right? Isn't that what the experts say whenever they are talking about ranking in organic search results? Usually, that means that marketers who are selling products need to write long, insightful content about their services or goods.

But that doesn't help much if your goods are the content itself. If your content is your product, you aren't necessarily going to win in the Search Engine Result Pages (SERPs).

In this white paper, we will talk about the most common organic search challenges facing publishers, and what some of the solutions are for each of them. These will include:



Possible keyword cannibalization for big terms



Archiving outdated content, while maintaining editorial standards and best practices



Important content that can be hidden behind JavaScript



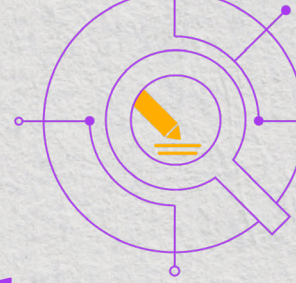
Getting news articles surfaced in Google News and hot news topic search results



Google's ever-changing recommendations around paywalls



Scalability across a large website and various types of content



More Content? More Problems!

Common SEO Problems for Publishers

Publishing websites host a high volume of content, with a large amount of new content produced every single day. The volume and pace of editorial decisions alone can pose a few challenges, but the website structure and design might create complications, as well.

Multiple Content Types

Most publishing sites aren't producing just one kind of content. Even a news organization will have reporting, editorials, FAQs, explainers, rich content such as videos or infographics, interactive maps, and so on. This is a management challenge from a CMS and analytics standpoint, but it also poses SEO complexities.

SEOs don't just have one set of SERPs to monitor, target and optimize for, but several. They might be watching Google News, Google Discover, AMP pages, evergreen pages, knowledge panels, image carousels, video carousels, and more.

Archives

News publishing sites, in particular, are challenged with keeping their archives clearly labeled and maintained. It's not always possible to link to updates on an ever-changing story. Sometimes older versions of a story have details that might be useful to someone doing a very specific long-tail search.

👉 On the other hand, it can be useful to dust off still-popular archives long after they age out and turn them into evergreen content. Managing content at scale like this can be tricky.

Topic Cannibalization

With all of that going on - and a lot of people working on things concurrently, it's only a matter of time before two pieces of content are produced on extremely similar topics with overlapping target keyword phrases. If the team is optimizing all the time (and I'm sure they are), that means you've got two content pieces battling for superiority in the search rankings.

While that sounds like a good idea in theory, in practice, it means they both hold each other back. Google only wants to rank one piece of content per publisher per topic. So if you have two pieces and it can't tell which one should rank, it's going to put both of them out there - but a few ranks lower than you usually would appear for that topic.

Two pieces of content go to battle in the search results, and only one can come out alive. Because of the amount of content on large publisher websites, you may not be aware of this issue until there's an unexpected drop in search position - losing your brand its hard-earned SERP until you can fix it.

Crawl Rate and Crawl Budget

Fresh, timely articles - especially news! - need to be discovered, crawled, and indexed ASAP. Sometimes scoops depend upon how quickly Googlebot can find your news.

What determines the frequency at which Google returns to your site, how long it spends on the site while it's there, and how many clicks deep it goes is called "crawl budget." This is the "budget" of how much time Google will spend its resources accessing and interpreting each website.

With millions of pages, a publishing website can run into problems with crawl budgets. Information architecture, page speed, and other factors all contribute to how frequently Googlebot crawls a website.

JavaScript Implementation

Because the crawl budget is critical to Google finding your extensive content, it's likely that page speed is too. That often means implementations using JavaScript (JS). JS is a blessing and a curse to SEO professionals. Important on-page content can be hidden behind a JS tag, it can often slow down pages or page elements, and updates for SEO purposes are normally backlogged with most development teams. At the same time, it can help with some of the various revenue model implementations for publishing sites and often allow for multiple methods of monetization

👉 It's a good idea to review the Google Search Console page inspection, and the various JS reports provided there for various content types and site sections to ensure that the JS isn't getting in the way of great search rankings.

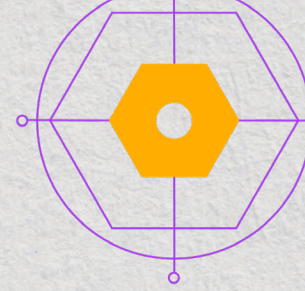
Things That Help Revenue Can Potentially Harm SEO

Sometimes the things like JS interstitials that require subscribers to log in can hide key content from Googlebot. Then, when the page is indexed, search engines don't have any information to go off of to rank the page. This can potentially lead to thousands of your pages not being ranked because there isn't enough on-page content for Google to evaluate quality.

Sponsored articles that are paid placements of content can also be a threat to organic search results. It should be very clearly labeled and policies should be in place to protect search users from advertorial content, in addition to any canonical URLs that are needed.

Ads can both slow down your pages and harm the user experience of a site, especially if they are intrusive. While those are ranking factors, the percentage of the page that is covered in ad space is also a consideration for organic rankings.

Now that you've got a clear picture of the SEO challenges for publishing websites, let's get into the priorities and solutions.



Publisher site structure

This section will break down publisher websites by page type, as well as address the key areas for optimization, possible schema markup to apply, and other important things to know.

We will also talk about information architecture and the pages that are crucial to surface to search bots, as well as dig deeper into JavaScript and how it works with Flexible Sampling best practices.

Types of Pages, Markup, and Rankings

Each of the different kinds of pages have areas of opportunity to rank and appear in search results in unique ways. There are often different fields that are useful to consider across each content type as you optimize.



Speed to Publish... and Value

A publisher's website structure has implications on how quickly and easily they can get new content launched and, in turn, discovered and read. The WordPress VIP platform is designed to be approachable and inclusive, easily integrating with existing processes and technologies so everyone - and everything - that touches your digital experiences can readily contribute with little to no points of friction.

The intuitive and flexible WordPress block editor brings a seamless drag and drop interface to publishers, with a robust REST API and deep library of available integrations to empower engineers, experts, and editors to get more out of their content management system.

In general, no matter what kind of page or content we're talking about, best practices would include schema microdata markup for your breadcrumbs - and ensuring that those breadcrumbs path a user around your site as easily and efficiently as possible.

👉 If possible, breadcrumb paths and URL structures should mirror one another for maximum UX gain.

Look, you've got this down. Your SEO team has trained the whole editorial staff on writing whiz-bang page titles and excellent meta descriptions and the designers know all of the image alt-tag best practices (and they know that what you name an image is even more important than the alt tag for SEO). Right? Right. So all those best practices are in place all day, every day. (Except maybe some of that old stuff in the archives...)

WordPress VIP customer News UK improved time to publish an article by 60% after making the move to the "Gutenberg" block editor, with a 30% reduction in the number of clicks it takes to produce a story.

"With Gutenberg, we were able to publish a breaking news story in two minutes versus five minutes," said Joel Davies, Head of Editorial Operations at News UK. "The main reason for this is the reusable blocks which have been renamed 'The Game Changer.' "

[Learn more.](#)

WordPress ships with the block editor, libraries of common blocks and layouts, taxonomies, topic pages, and plenty of opportunities to add custom reusable blocks and templates.

Enterprise publishers use the WordPress VIP platform to layer on scalability and performance optimizations that keep their sites running smoothly at the biggest moments for their business.

Articles

Whether an article is an op-ed, an evergreen explainer, or a general how-to, you've got content on a publishing website that is just a run-of-the-mill standard article.

Key areas of an article to optimize:

- Page Title and Meta Description
- Titles and subtitles
- Headers
- Content and images
- Canonical URL (if necessary)
- Breadcrumbs and breadcrumb markup
- Related links
- Source boxes (if necessary)
- Bylines
- FAQ markup (if applicable)

👉 One of the most overlooked aspects of articles is how they fit into your information architecture. Often, publishers think about leading the Googlebot to a new article and how the architecture of the website does that. What they don't think about is that that same article can be a landing page for a given topic, and the related links in the navigation and contextual links in the body copy can be incredibly valuable for creating a picture of all of the content your site has on a given topic.

Source boxes are fabulous page elements to tag onto the back end of articles if you've had to do a lot of research to write an article. They are useful for two key reasons:

- It helps you know whether an evergreen article is still up-to-date
- It boosts your authority on a topic, signaling that you can be trusted

👉 You may have heard of E-A-T signals in SEO. That stands for Expertise, Authority, and Trustworthiness. That source box at the end of an article is an impactful way to show a Quality Reviewer that you are trustworthy!

A byline with a link to a personal profile can also be helpful for E-A-T. A person's individual authority can also confer authority to a page. This is where the schema "person" markup can come in handy.

Even if your publishing site is primarily generating news articles, it's a good idea to have a stock of evergreen pieces in your content mix. You might need the general "what is it" pieces that explain complex concepts that are newsworthy from time to time. Because news traffic is spiky - each article gets a lot of hits for a few days, then trails off to nothing - having longer-lived evergreen articles will take some of the volatility out of your traffic metrics.

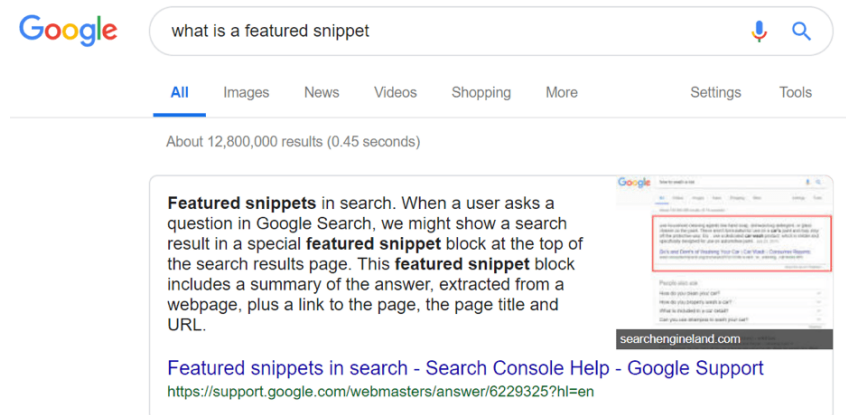
Article SERP Features:

The key places in search engine result pages where article content can get highlighted by Google are:

- People Also Ask



- Featured Snippets



Above, we mentioned FAQ markup. If the evergreen content in question is an FAQ, schema FAQ markup is a quick, easy way to get placement in the People Also Ask box.

Unfortunately, there isn't a huge traffic win from that placement (Google users tend to get the answer they are looking for right on the page), but you win a little brand recognition and you take over more real estate on the result page.

The Featured Snippets block is the box at the top of search results that often answers most of the questions that Google thinks a user is looking to answer.

👉 That means that a Featured Snippet box can give you insight into what Google thinks the user's intent is behind a given search phrase!

It also means that if you can answer that question succinctly and clearly, you have a shot at appearing in that box. Generally speaking, the website that appears in the Featured Snippet would usually organically rank in positions 1-5 on the first page of Google no matter what. What the Featured Snippet placement can do is take a website that ranks #5 on the page and skyrocket them up to that position zero above the organic results.

Just like the People Also Ask, often search users get what they are looking for from the content that appears in the Featured Snippet box. This is why the url in the box has a slightly lower CTR from Google than the top organic ranking.

How do you get placement in a Featured Snippet, you ask?

👉 Best practices for Featured Snippet placement are:

- Include the question you want to answer in sentence form in an H1 or H2 on the page.
- Answer the question as succinctly as possible immediately below that. If you can do it in one or two sentences, that's ideal. If you can do so using a bulleted or ordered list, that's even better.
- If it makes sense for the content, have the header above your answer in question form (ex. What Is Evergreen Content?).

Sponsored Articles / Advertorials

If your publishing site produces advertorials or sponsored content, you must make sure that any links pointing away from those pages are tagged with a nofollow robots directive.

Full disclosure: the other thing that's important about sponsored content is to not be sneaky about it. It should be clearly marked as sponsored and tied to the paying party in the visuals.

While nofollowing links are an absolute must, you don't have to mark sponsored articles noindex, but you still might want to if they compete with your non-sponsored articles on the same topic.

News Articles

Google News is a special animal. It uses a different index, a different series of signals, and it clusters and arranges results by topic in a different way than the main index does. The news team has described their topic clustering pattern to be more like a blob or amoeba, where it grows until it has to split off into multiple topics. Capturing a top spot in those blobs is the goal for a news article type.

News articles get all of the same optimization sections as the evergreen articles above, and a few that are special just for News articles.

- Author markup / Person markup
- Dates

News articles also have their own special cul-de-sac in the Google webmaster suite of tools. In the [publisher center](#), news publishers can submit separate news xml sitemaps, and management of those pages can happen without the rest of the Google Search Console tool. Of course, dates are super important when it comes to news articles. One of the best practices (though no longer a requirement) is to have a date in the URL of news articles. However, even more important than that is including one clear, consistent date at the top of the page for your users, and to use the date-related schema markup:

- datePublished
- dateModified

Google's guidelines warn against having artificially refreshed dates or deceptive update practices. Be sure that all of these dates align with reality, and you'll put your best foot forward.

Using person markup for authors is no longer supported per se, but as mentioned with the bylines above, having articles linked to specific writers or experts can signal trustworthiness overall.

Google has deprecated a lot of the tools it gave news publishers over the years (such as Editor's Picks or the standout tag), but there's a Google News Initiative that is still working toward highlighting and rewarding the best in class journalism that's out there on the web. In fact, there's a new program called "[Google News Showcase](#)" being tested in Germany, Brazil and Australia that is supposed to roll out more in the near future.

👉 Content duplication is a huge risk with news articles. If you're not ranking well in the news topics, consider being very aggressive with canonicalization and cutting down on any possible duplication that's happening on and off your site.

AMP Articles

Because mobile friendliness is an absolute must for news articles as well as page speed. News is frequently the content type that is most often targeted to be converted into "AMP" pages.

AMP stands for "Accelerated Mobile Page" and is a Google product intended to speed up the mobile web. Pages are simplified to the barest of bones, rewritten in an open source code called "AMPhtml" and cached on Google's cloud of servers, rather than your site's host. You have markup in the source code that pairs the standard page to the AMP page similar to an HREFLANG that "handshakes" between the two sets of articles.



Simplifying AMP Readiness

WordPress is used by nearly 40% of all known websites today, according to industry trackers. The ubiquity of WordPress means when Google, Facebook, Apple, and other large technology companies build new features for publishers, WordPress is often an integration partner at launch.

The AMP Project is no different. The [WordPress integration for AMP](#), developed by engineers at Google, WordPress VIP, VIP Agency Partner XWP, and others within Automattic and its ecosystem, supports efforts such as generating AMP-valid markup and the

It's a bit of a challenge to set up, but once you get the tech part established, it can be a great way to appear higher in mobile web search results. You also get a graphical element to signal to users that it's an AMP page, but there's no data showing that users care too much about that little highlight.



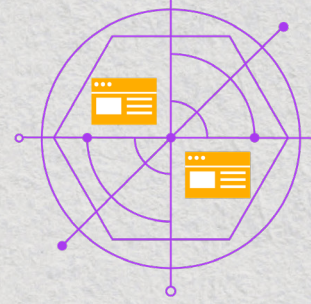
...serving of AMP pages. Working with the AMP Project team, we've helped to maintain and support the WordPress integration for enterprise publishers.

AMP, [Google Web Stories for WordPress](#), and Core Web Vitals are just a few of the projects where WordPress VIP represents our enterprise publishing customers, along with partners like Botify.

On the other side, there's a good bit of maintenance. Google also only passes the barest of analytics data to you from their server, so you may not have a clear picture of who saw your AMP pages. They can also sidestep a good bit of publisher revenue models, so be careful with how ads and paywalls are treated if you consider implementing them.

The other major downside with AMP is that they basically double your site's size, and if you have crawl budget issues, AMP can exacerbate them.

Publishers have great success with AMP, and as many have chosen to optimize their mobile sites for page speed on their own servers. AMP isn't a "get out of page two" free card, so implementing it doesn't guarantee rankings.



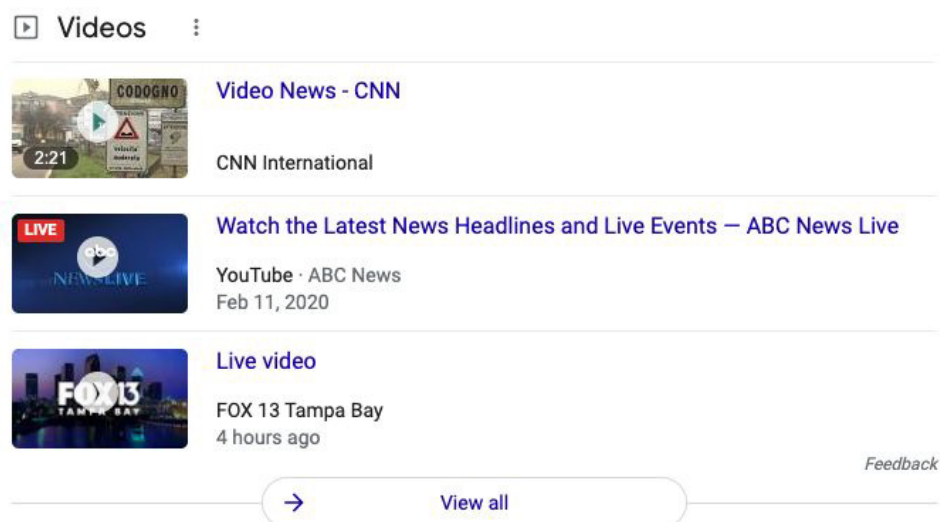
Rich Content Types

Most publishing sites have more kinds of content than plain text articles! Videos, galleries and slideshows, quizzes, recipes - the list goes on (and on)!

Naturally, each of these content types have their own opportunities for optimization. And just to complicate matters for the SEO team, most of them also have their own SERP result features to target.

Video

Video rich snippets show the search result with a thumbnail image of a video instead of a plain text link. This often appears where a search query might best be answered via a video, instead of written content. The way to get into these is video markup.



Video packs and carousels are another way Google highlights video content on the SERPs. This is often 3-5 video results with thumbnails. You don't have to use video markup to get into the carousel, but it does help. The other way is to just upload that video to YouTube (since that's a Google owned property).

If a publishing website is also a broadcasting resource, then there is an additional video feature - a LIVE badge. This is to show streaming / live coverage of news events, sporting events, or award shows.

Videos, articles and recipes are all eligible for How-To graphs. These have special markup breaking out the steps in how-to content and presenting them in a rich way in the SERPs.

How to Write a Book in 15 Amazingly Simple Steps

1. Find your "big idea" The one thing you absolutely need to **write a book** is, of course, an idea. ...
2. Research your genre. Once you've found your big idea, the next step is to research your genre. ...
3. Create an outline. ...
4. Start off strong. ...
5. Focus on substance. ...
6. **Write** "reader-first" ...
7. Set word count goals. ...
8. Establish a healthy routine.

[More items...](#) • Jan 28, 2021

[blog.reedsy.com](#) > [how-to-write-a-book](#) ⋮

How to Write a Book in 15 Amazingly Simple Steps

[About featured snippets](#) • [Feedback](#)

Some of the markup is required in order to give Google enough information to know what to do with the asset. Here are the fields that are required:

- VideoObject
- Name
- ThumbnailURL
- Either ContentURL or EmbedURL
- Description

The rest of the markup is optional. If you have the data, it's better to include it. And some of it has bonus perks, like the LIVE badge.

- BroadcastEvent (More details from Google are available here)
- Clip (a defined 30 second or less segment of video that targets a specific topic)
- Duration
- UploadDate
- Expires
- Height
- Width
- HowTo
 - Name
 - Step
 - Description
 - (then you can drill down to [specific timestamps in the video](#) per step!)

If your site has an index or list of videos, and you'd like to be featured in a Video Host Carousel (where the pack only contains your videos!), the way to do this is to apply the **ItemList** markup to each marked up video on a gallery page.

Aside from schema, there are a few more best practices to getting video content to rank.

👉 One of the more common ways for videos to hide from search bots is to put the player in JavaScript along with the transcript of the contents. Ideally, the transcript would appear on the body copy of the page or the html of the page. If that's not feasible, it should be available in the <noscript> tag. This is important for usability and accessibility as much as for robots.

Additionally, please don't autoplay videos when users land on the page. This can make for a poor user experience, which can affect your SERP position if most users immediately leave the page when they arrive.

Galleries and Slideshows

Photo galleries and slideshows are the bread and butter of websites with ad and pageview bound revenue models. The good news is that they can be really popular with users, and if built properly, can rank pretty well in organic search. The bad news is they can be hogs on page speed, and the content has to be worth the user's time.

First, let's talk about image optimization. There are three ways to improve organic search with images: file size, text descriptions, and schema markup.

File size is the biggest of these because your images can be the biggest culprit of page speed problems site-wide, not just on gallery pages. Some publishers are handy about resizing the images to compress them down to that sweet spot between high-quality and lower size. Depending on your CMS, there are numerous plugins that can assist with this as well. Many publishers are moving toward the newer file compression formats. The next-gen formats are great for resolution and page speed, but they don't work across all browsers. It's important to know who your users are, and whether that impacts the bottom line or not.

There are several fields available to describe images on your site via text. Yes, Google has a machine-learning reader that's been working on interpreting image content for several years, and it's getting pretty adept at it. But that doesn't mean it's fool-proof, or that you want to give Google's guesswork the control over describing your valuable images.

HTML elements where you can add text and context about an image include:

- The image filename
- Alt text
- Captions
- Title
- Description

All of these fields are available in standard markup. SEO tests show that the file name on the image and the alt text are consistently the most important. If text is inside the same div block as an image - such as a caption would be - it's also pretty important.

👉 Alt-text is not intended to be laden with a paragraph of search terms. It's used for screen readers, and should be as clear and succinct as possible, while being very descriptive of the image.

Finally, there are two different options for schema markup on images. Adding and identifying images are easy - in terms of the required fields.

- ImageObject
- ID
- ImageURL

The optional fields for images is where it gets interesting! All of the options in the [ImageObject](#) type are available. This includes the height and width dimensions of the image, the data created, or the photographer's name. (Note: [Logos have a separate type of markup](#) that includes the ImageObject information, and a few extra fields.)

For stock photo providers, there's a [Licensable](#) tag to help market images that can be purchased. For publishers of original content, that license markup can help provide claim of originality and ownership.

[ImageGallery](#) is markup that helps explain to robots that there are a series of images strung together, even if they don't all appear on the same page. This is key for those galleries and slideshows! This allows you to list all of the relevant imageObject markup on the first page of the gallery, even if the thumbnails for those images aren't available on that first URL. This is also a faster way to load that initial url than including the image objects in the initial page load.

The other way to manage this view of all of the images and their relevant content is to enable a "see all," and allow Googlebot to crawl and index that page as well. The danger there, of course, is that search users might be directed to that page as a better experience, and then those ad impressions and pageviews are null and void. It all depends upon how the content is monetized.

Last but not least are the images themselves. Google image search can be its own organic search driver. Check out the image search data in Google Search Console and see how your images are performing all on their own. Ranking in the top five placements of image search often means that your image might be pulled into the image carousel that often appears on the first page of organic results.

Stock photos will likely only rank for websites selling stock photos - because they are duplicated and not necessarily extremely relevant to the content on the page. The more closely related the image is to the content on the landing page, the likelier they will rank.

Unique, proprietary, original photos, graphics, infographics, and other media are far more likely to rank than stock imagery is. Again, make sure the text on the page provides plenty of context for the chart or infographic for usability purposes.

Quizzes

In the publishing world, quiz content tends to be more targeted to social media audiences than it does organic search. However, that doesn't mean they can't rank and perform well when properly optimized.

The key thing to do here is expectation-set for a search user. Make sure the title and meta description are clear about the fact that the user will be landing on a quiz. It will decrease the click through rate for the asset from the search result page. By setting accurate expectations, you'll have a lower bounce rate, and happier, more engaged users.

On the page itself, consider adding ImageGallery markup if images appear throughout the quiz, and ensure that the quiz has mobile-optimized interactions and page speed.

Quizzes are often JS driven, so the important part for SEO is to have the first page and the last page visible in html or `<noscript>`. If you can reveal the questions to the bot, that can provide more contextual relevance for rankings.

Further, because quizzes and quiz results are often talked about by users they can be bait for inbound links, which is important to SEO. Since the result page is usually the thing that gets linked to, that's the page you want to link from.

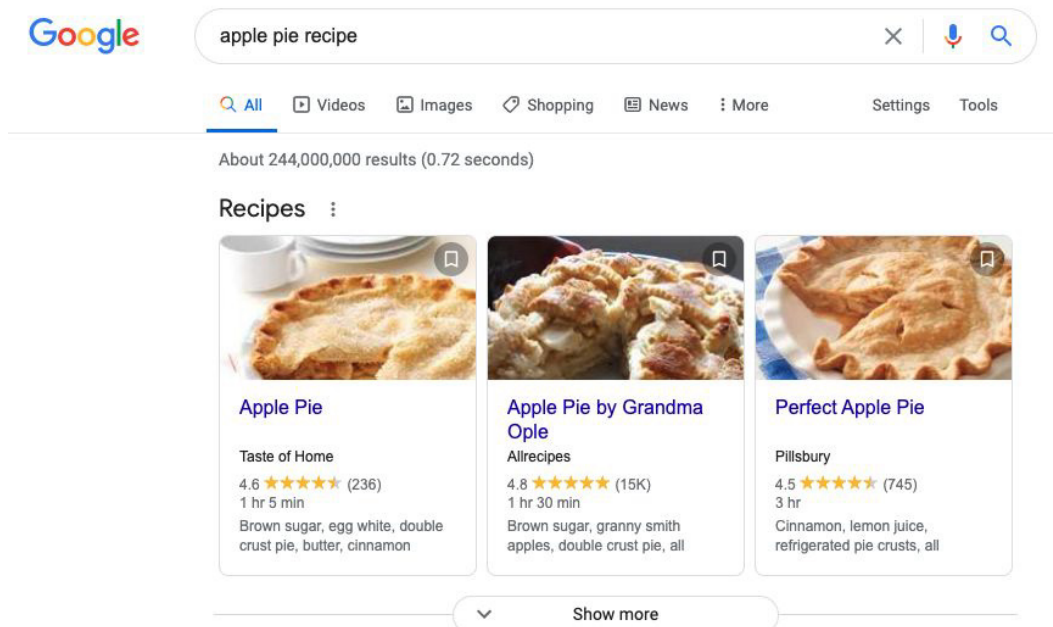
👉 Be sure to add a few highly contextually relevant links at the bottom of that quiz result box. Those internal links will benefit from the boost!

Recipes

Recipes in organic search results are so compelling to a search user that markup is pretty much required to rank and compete.

Recipe rich snippets are really informative to a search user. They include images, reviews and star-ratings, cook-time and special dietary information all right there in the search results page.

If a person Googles a general query, like “apple pie recipe,” there will be four pages of rich results before the plain text entries appear. Because of this, the more specific your recipe pages and their associated markup can be, the better.



Required recipe markup is very minimal. It includes:

- Image
- Name

If you want the Google assistant to help walk you through marking up your recipe content, that list expands to:

- recipeIngredient
- recipeInstructions
 - HowToSection
 - HowToStep

Further recommended fields include:

- aggregateRating
- Author
- cookTime
- prepTime
- totalTime
- datePublished
- Description
- Keywords
- recipeCategory (is it a meal, dinner, snack?)
- recipeCuisine
- recipeYield is necessary if you want to include
 - Nutrition.calories (or any other nutritional data)

If the recipe includes a video, you can also use VideoObject markup to have the video thumbnail appear in the rich result instead of an image thumbnail.

Similar to videos, using an ItemList markup on a gallery page of a number of recipes can enable a host carousel of your recipes on branded search results.

Naturally, just like articles, the page title, H1 on the page, meta description and subheadings on the page still matter in recipes. Because the competition is steep, the page title might be more important than it is in other results.

👉 Testing page titles for CTR improvements is a good idea with recipe content if you've got the rich results appearing, but traffic still isn't flowing through.

Infographics

Infographics and other forms of data visualization are so important to helping people understand complex topics. Optimizing data visuals and charts is very similar to images, but not quite the same.

Of course the file size, page load and compression are all key. Also, the same fields identified above in both html and schema markup are important.

Additionally, infographics and data visualizations aren't accessible to visually impaired users. If you take the time and effort to really describe your chart and graphic to a visually impaired user, and to present that in the code of the page - or better yet in the content of the webpage - what you've done is a win-win. First, you've improved accessibility. Hooray! Second, you've given a detailed context clue to the Googlebot about what to do with your graphic.

Like quizzes, graphic assets tend to do particularly well in social media, and can act as link bait.

👉 If you have content in the body of the page that describes the graphic, anchoring some internal links from that content can pass some lovely relevant, contextual link flow to target pages on the same topic.

Interactive Assets (Maps, calculators, etc)

Now that the world is online, we all realize that interactive pages are what the internet is all about - no one expects plain text to sit on a page anymore. We want to click, comment, expand, or adjust the world to our liking.

Ranking for interactive assets sort of follows the same guidelines as infographics or quizzes, but with a twist.

It's going to sound redundant to bring up page speed and mobile usability here. But it has to be mentioned.

Also it's going to sound pretty repetitive to discuss JS.

This is where the `<noscript>` tag is going to be most useful to ensure that Google understands the functionality of the tool on the page.

👉 Be sure that Googlebot can access and crawl your JS on the page of the interactive. You want them to be able to render the tool completely.

Like quizzes, setting search users' expectations about the kind of interaction they are getting when they click on the link is the best way to have low bounce rates and high engagement with interactive tools.

If there's a results page, or a shareable element, those internal links are a really good idea of something that might get a lot of links from around the web.

Community and Forums

For many publishing websites, user-generated content (UGC) is a crucial part of the content strategy. Users will write entire websites, and communities and forums can create a support system for users that the site never thought possible with limited resources.

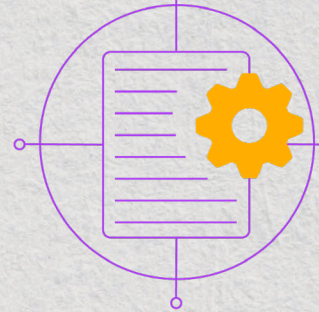
On the other hand, there's a huge amount of upkeep necessary with forums. To use UGC to manage super niche, long-tail content, here are a few things to keep in mind:

1. Each new question/topic should create its own unique landing page with a specific url and unique page title (using the words from the question/topic automatically).
2. Topic categories and lists should have their own unique landing pages with a targeted, SEO optimized title, meta description, URL, and preferably a few lines of body copy setting the stage for the category
3. Try not to set up too many categories at the beginning. It's important to have something on these landing pages and have them be rich and useful to readers, not just 1-2 questions.
4. Assign admins per category who have power to block / boot problematic users, delete spam and moderate bad behavior.
5. Definitely allow images to be uploaded, but include file size restrictions!
6. Consider using the robots.txt crawl exclude function to limit Google's visibility of user profiles, logged-in states, etc.
7. If you expose user profiles to Google, allow the users to opt-in or out of being indexed. Give them a radio button to control their status of having a meta robots "noindex" tag in their profile page or not.
8. Be sure to nofollow any links that come from this portion of the website
9. Set up enough controls to prevent spam robots from taking over your pages

If that sounds like a lot of work, it is. But the payoff can be huge in terms of SEO traffic. In addition to allowing you to help users find the answers to very specific questions that you'd never have the budget to write content about, it also allows you to take the pulse of your community.

Logins also allow you to create a growing list of emails and social accounts that you can tap into when you need to drive traffic to a particular topic.

Finally, you can mine that community for topics on which you might not have known you wanted to write evergreen content.



Crucial Pages for Indexing on Publisher Sites

We've already teed up the idea that publisher websites need to optimize the amount of time that Googlebot spends on the site and where they spend that time. That means that you might want to prioritize key areas of the site for crawling and indexing - and that there might be areas of the site you don't want the bot to bother with.

👉 Remember, if a page isn't linked to from your site, Googlebot can't discover it.

In fact, the bot usually starts at the homepage, and follows all of the links from there. Each time it clicks on a new link, SEOs refer to that as a "level" of crawl depth. You want to ensure that the newest and most important links are not too "deep," or it will take a long time for them to be discovered and updated in the index.

Key Areas to Index

Crawl frequency is how often the Googlebot returns to certain pages of your site to check for updated information. The most important pages should be crawled relatively often, and they need to be close enough to the homepage to warrant that level of attention.

While all websites vary, the key areas to index in a publisher site might include the following:

- News publication pages
- News articles
- News archives
- AMP pages
- Evergreen content

If a publisher produces content targeting Google News and Google Discover inclusion, either the main news page or the homepage will be listed in the Publisher Center to ensure regular check-ins from the bots.

This page should include automated links to every single news story produced in chronological order. It might have excerpts from the top headlines, or some contextual content to ground a possible user arriving to the page with an understanding of what they are looking at. Ideally, this page is what the News item in the breadcrumb in every news article points to.

Naturally, the whole point of this page is to get the news articles themselves crawled, cached, and added to the index. These links to the new articles are ideally anchored on relevant text (as opposed to contextually irrelevant prompts like “read more” or “click here”), and use the canonical (or official) URL for the news article.

Generally speaking, this news page is also paginated to the beginning of time. That means that eventually, the pages of links include the entire news archive in chronological order. This is important for that long-tail search traffic mentioned above regarding news archives.

It’s okay to mark the deeper pages of this list of news links “dofollow,noindex” - that tells the bot that those aren’t intended to be landing pages for search users, but that they are useful for the Googlebot to keep checking out.

The AMP pages associated with these articles will be listed in the source code for each article. So the indexing should have a cascading effect, where first the news page is crawled, the articles are indexed, then in their code, the AMP pages are discovered, crawled and indexed as well. I say “should” here, because that’s how it’s supposed to work. You may need tools to double-check that it’s actually working that way as many publishers find this to be finicky. It can be useful to include AMP pages in sitemaps. If you’re having trouble with AMP pages not being indexed, you might want to create a separate, specific sitemap for that content to better understand the errors and details. Google Search Console does do a lot of reporting around AMP because they want webmasters to adopt it.

No matter what kind of publisher you are, you want Google to crawl, index, and access your evergreen content.

The best way to do this is to provide a pretty clear information architecture with cascading breadcrumbs that make sense in terms of categories and granularity. The landing pages for the breadcrumbs are a lot like the news landing page described above - an updated list of links with a content blurb to ground a user. Again, a “dofollow, noindex” is ideal on this kind of page. For evergreen content, instead of chronological listings by publication date, you might want to do the chronology by update date or modification date.

Restrictions Via Robots.txt

The way that robots.txt restricts the Googlebot is by “crawl exclusion” - it says “don’t crawl this.” The crawl exclusion means that the bot won’t go there at all. The robots.txt file should be complete without being overly aggressive.

Things that are typically excluded are:

- User logins
- Session IDs
- Tag pages
- Shopping cart and payment pages

Crawl exclusions are great ways to reduce how much you’re asking the Googlebot to crawl without impacting search traffic. However, it’s really easy to crawl exclude too much, so be careful to allow access to JS files, images, and other key areas of the code.

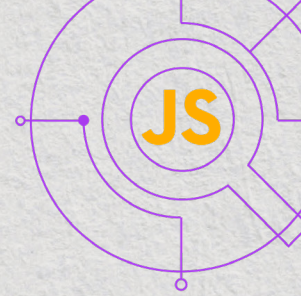
Finally, SEO best practice is to include a link to the xml sitemap in the robots.txt. It’s a nice tidy way to keep all of the bot-speak together, but honestly, if you have sitemaps in the html and in Search Console, it’s not really necessary.

Leveraging Noindex Tags

Finally, the other way you can tell the Googlebot to deprioritize areas is to signal that they should not be included in the index. A meta robot's "noindex" tag is an important tool to signal to search engines that a page is not a suitable landing page for search users. Google will still crawl the page (though it appears they crawl noindexed pages a little less frequently), but they won't include it in search results.

This is useful for:

- Pagination
- Tag pages
- Pages with little to no content or duplicative content
- Pages that are useful to users, but not good landing pages for someone starting out on the site



JavaScript and Flexible Sampling

Google can crawl, execute, and render JavaScript (JS), and has been able to do so for several years. However, JS execution is resource-intensive. While Googlebot can see this information, it stores the data until it has a reason to render the page more fully. More often, the text-only, mobile-only bot is likely to be what is skimming the page - unless it sees big changes from the last render.

When JS rendering causes page speed to slow down, the Googlebot is even less likely to allocate the resources required to wait for a page to load. This is because a user is also unwilling to wait around. JS can be a mixed blessing in terms of page speed - it can be a huge improvement, or if done improperly, a cause for slow page loading times.

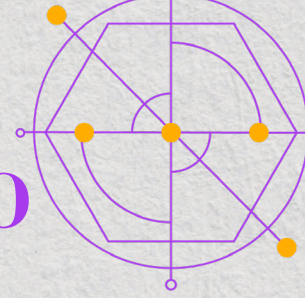
Because it renders the information in JS less frequently, links and content that appear exclusively within JavaScript elements appear to be given less value. This is something to be aware of when deciding upon technical implementation:

- Elements designed to navigate and break up content. “Tabbed” designs are common in more technical forms of information - where they look like mini tabbed menus on the page. Sometimes these are database driven sorts of site sections. It’s possible to reveal all of the content across all of the tabs in the html while still using JS on the tabbed experience. This is similar to a lot of pagination and “read more” kinds of implementations intended to allow only a small part of the page to load initially, and requiring the user to interact to see the rest. Again, this is good for page speed and sometimes for usability, but you can hide so much information behind JS if it’s not implemented correctly.
- Roll-over or mouse-over definitions or mini paragraphs of text often called “tool tips” are another source of great content on the page that can be hidden in JS. Again, it’s possible to reveal these in the html rather than keeping them all in the DOM.

- “If you like this article, also read this...” is a key way to get web users to spend more time on a site and engage with more content. Internal links to recommended articles are often generated automatically in a JS module and dynamically personalized for users. While this is fabulous for UX and often for conversion, these links don’t pass a lot of SEO value. There’s often a need to create a hybrid approach for internal links. Links that are in the body copy in the html are higher value.
- Ads and sponsored articles are often served via JS because they personalize the ads to specific users, or automate ad serving to ensure that goals are met. This is a great way to use JS on the page, because it’s also stuff you don’t necessarily want the Googlebot to pay a lot of attention to. The only thing to watch out for here is that JS should be optimized and minified so it doesn’t slow down the page.
- User-Generated Content - such as article comments and recipe reviews - are extremely valuable parts of a page that are frequently included in JS elements. By having this information loaded dynamically, it’s possible that the unique, rich, relevant content found there is not surfaced on those pages. There are various ways to solve this, and SEOs have to know which elements on the article pages are crawled less frequently due to JS implementations.
- One of the most common ways for publisher websites to use JS is to choose which content to show before a paywall or what content hides behind it. We’ll get into revenue models below, but this is probably the most important place to have visibility in terms of what is visible and what is not - both to users and the Googlebot. The dynamic flexibility of JS is ideal for this purpose, and with their [flexible sampling](#) guidelines, Google supports it.

Don’t forget the `<no script>` tag and other important text that should be available on the page. Alt text for images, descriptions, and video transcripts are all important elements that should be surfaced to the text-only bot every time it visits.

👉 Be careful with using robots.txt to restrict the Googlebot’s visibility in js folders and files. While it’s a great way to manage ad-server content, it’s not a good idea to limit their visibility of valuable content!



Publishing Revenue Models and SEO

Generally speaking, publishing websites are monetized through sponsorships, ad impressions and clicks, and on paid subscriptions. None of these is particularly SEO-friendly, but there are ways that SEO and revenue models can work together.

SEO delivers the ad impressions, to a certain extent. A search visitor tends to spend less time on the site, and is less engaged than a user clicking through from a newsletter or social media. On the other hand, there are often a lot more search visitors.

The trick with ad revenue is that it lures the coffers to be intrusive and obstructive with ads. Inserting ads in the middle of body copy, having more ads above the fold, autoplaying preroll on videos, and having ads fly in and obstruct the page are many ways that ads are being served on websites that depend upon impressions and clicks on ads. And most of those things are negative search signals.

There is a fine balance between battling ad-blindness and harming search rankings. It's important for an SEO to have an awareness of what the ad servers are doing on every page to be able to tell where things are causing problems.

We've already discussed affiliates and sponsored articles. The key to those is nofollow tagging everything. Like the ads, they create risk for search rankings, and require visibility.

Finally, the most common way that publishers ensure revenue is to add a paywall, and get users to pay for a subscription. Often, they do this by cookie-ing users and restricting their ability to read articles after the first five or ten pages.

For years, Google demanded a first-click-free model, where the first time a user from a search engine clicked on the page, a pay-way could not appear. This wasn't always how websites wanted to gate or display their content. In recent years, Google moved to a flexible sampling model, where they encourage publishers to use the CSS to determine what content is being displayed to users or not. They don't prefer the use of JS for this application, but they can work with it.

The way to use JS and schema together to surface paywall content in Google is the Subscription markup. Applied to any of the CreativeWork types of content (article, news article, video, etc), "IsAccessibleForFree" can be set to FALSE in the schema markup, thereby indicating that it is subscription-only content. Usually this is indicated in the CSS with a div class of "paywall" or something similar to make it exceptionally clear.

When using flexible sampling, the content should be surfaced via the DOM to the Googlebot, even if it's not in the html, so that they can still rank and index it.

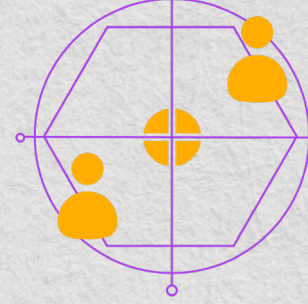


Revenue Model Experimentation and Implementation

Enterprise publishers have strong requirements to diversify their revenue models, with subscriptions, memberships, and a variety of paywall-style systems top of mind. And with partners including Piano and Zephr, WordPress VIP offers the necessary platform flexibility, supporting many proven subscription models and experiments.

Of course, when implementing or testing new models, however, it's important to also consider and monitor the impact on content accessibility, search traffic, and website performance. In fact, when WordPress VIP customer Observer Media was building out a new subscription offering, they conducted substantial user research, thought through a number of business implications, as well evaluated their partner ecosystem.

Following launch, the Observer's subscription features helped drive revenue ahead of their budgeted goals, while serving as a 4x multiplier for their registered free users -- all with no meaningful drop in traffic. [Learn more.](#)



Being a Team Player: SEO as Part of a Larger Team

For publishing websites, SEO is driving a large portion of the traffic that is driving revenue.

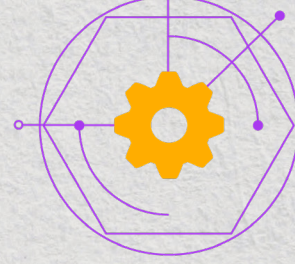
However, SEO is not the only part of this puzzle. UX, Marketing, Sales, Editorial, Design and Engineering are all doing their parts in the larger ecosystem of the website.

👉 Everyone who can impact the content, display or function of the website is “doing SEO” whether they realize it or not.

It’s important for the variety of teams to understand SEO best practices where it applies to them. A few great ways to promote this include:

- Cross-functional and cross-departmental SEO training and Best Practices
- Google updates via “lunch and learns” to help everyone feel smarter about SEO
- Access to tools and reporting with key stakeholders
- Provide reports showing the success metrics for cross-team projects

There is a common thread to these ideas - because the SEO team is often one of the areas that is looking at the whole site as a functional ecosystem, there are often details and pieces of information that only the SEOs understand. This is a valuable resource that is made more valuable by sharing it across teams.



Prioritizing SEO Projects for Publishers

SEO Audits tend to be a huge long list of things that need to be improved on a website, without the additional context of how much work each thing takes, or how much return the company will see on the investment of time and effort to improve it.

Usually, the design, engineering and editorial resources that are required to improve the SEO on a site are shared across multiple business teams and have competing (and sometimes conflicting!) goals. SEO is in a unique position to help prioritize these.

Individual projects might include detailed metrics such as increasing CTR from Google, or increasing visibility in image carousels and traffic to product pages as a result. But the bottom line in all of these cases is that the goal of the business is to increase revenue. If possible, an “apples to apples” comparison of conversion and sales improvement by project can often help stakeholders determine which projects take priority.

Of course, if readers cannot find pages (e.g. they are not being indexed), then any increase is a significant increase. Another way to weigh priority would be the number of pages impacted by the project. For this reason crawl improvements can take priority over goals that more directly affect conversion.

Get the Information You Need to Make Decisions

Before any of these priorities can be set, however, the SEO team has to be able to see what is going on across the site and to assess potential problem areas. The team might know that it takes a while for product updates to get indexed, or for key pages to appear in the index at all, but they might not know why.

The Botify Publisher tools are here to help. <https://www.botify.com/solutions/publishers>



Crawl and render your JavaScript at 100 URLs/second. Quickly pinpoint slow-loading resources, how your content looks pre and post-rendering, and perform device-specific JavaScript analysis.



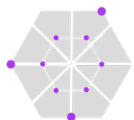
Optimize Crawl Budget. Direct search engines away from unimportant pages and to your key pages by using Botify to identify which pages they're crawling vs. missing.



Audit your structured data at scale. Monitor and audit the structured data applied across all of the various content types to ensure that you're targeting rich results.



Monitor your AMP pages. Gain visibility and ensure that the link handshakes are accurately maintained to ensure that AMP pages are working as expected.



Manage article archives. Use Botify to analyse which parts of archived articles are the most important and determine the best way to reuse and optimize old articles on the fly.



Looking for an easier way to identify SEO opportunities at scale?

Botify has you covered!

Botify makes it easy to save time, identify never-before-seen opportunities, and drive results — no matter how large your site!

[Request a Demo](#)

botify

101 Greenwich Street
New York
10006 NY

www.botify.com