/ theory / in / practice

"**Written by some of the top minds in SEO!**"
—Danny Sullivan, Editor-In-Chief, SearchEngineLand.com
& Producer, SMX: Search Marketing Expo

The Art of SEO

Mastering Search Engine Optimization

Eric Enge, Stephan Spencer,
and Jessie C. Stricchiola

4th Edition

Early Release

RAW & UNEDITED

The Art of SEO

FOURTH EDITION

Mastering Search Engine Optimization

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

Eric Enge, Stephan Spencer, and Jessie Stricchiola

The Art of SEO

by Eric Enge, Stephan Spencer, and Jessie Stricchiola

Revision History for the Early Release

- 2023-01-12: Sixth Release

- 2023-03-14: Seventh Release

See http://oreilly.com/catalog/errata.csp?isbn=9781098102616 for release details.

CHAPTER ONE

Search: Reflecting Consciousness and Connecting Commerce

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 1st chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the editor at sevans@oreilly.com.

Search has become integrated into the fabric of our society. More than 7.5 billion Google searches are performed per day, which equates to more than 85,000 queries per second.[1] In addition, users have grown to expect that the responses to their search queries will be returned in less than one second. Further, 53% of all website traffic across the web comes from organic (non-paid) search, which means that for many businesses, SEO is the most important digital marketing investment 2.

Through the power of search we're often able to find what we want in a minute or two, or even just a few seconds. People can use search to conduct many of their research, shopping, banking, and social transactions online—something that has changed the way our global population lives and interacts. As a result, it's critical for owners of web sites to increase their visibility in search engines as much as they can. However, obtaining such prime search result real estate is not a simple matter, but it is one that this book aims to deconstruct and demystify as we examine, explain, and explore the ever-changing art of search engine optimization (SEO).

The Mission of Search Engines

While those on the internet are free to use any of the many available search engines to find what they are seeking, Google remains the dominant player worldwide with more than ninety percent market share. 3 Nonetheless, the burden is on Google (and other search engines) to provide a relevant, fast, and fresh search experience. For the most part, search engines accomplish this by having the most relevant results and delivering them the fastest, as users will return to the search engine they believe will return the results they want in the least amount of time.
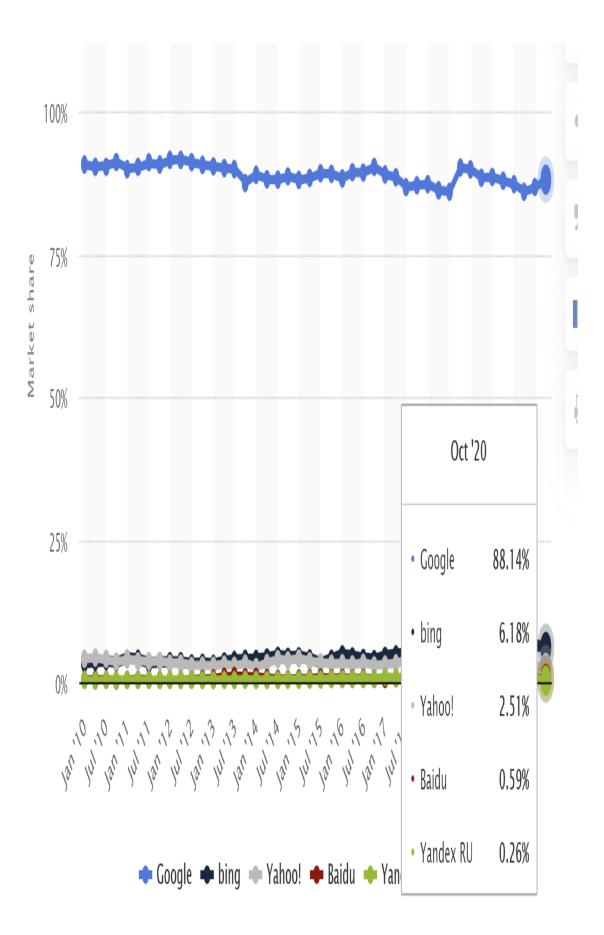
To meet this demand, search engines invest a tremendous amount of time, energy, and capital in improving their speed and relevance. This includes performing extensive studies of user responses to their search results, continuous testing and experimentation, analysis of user behavior within the search results (discussed later in this chapter), and application of advanced machine learning algorithms to tune their search algorithms.

Search engines, such as Google, generate revenue primarily through paid advertising. The great majority of this revenue comes from a *pay-per-click* (or *cost-per-click)* model, in which the advertisers pay only for users who click on their ads. Because the search engines' success depends so greatly on the relevance of their search results, manipulations of search engine rankings that result in nonrelevant results (generally referred to as *spam)* are dealt with very seriously.

Each major search engine employs teams of people who focus solely on finding and eliminating spam from their search results (generally referred to as "web spam" teams). In addition, larger search engines such as Google also apply dynamic algorithms that detect and deal with poor quality content and/or spam automatically. These efforts to fight spam matter to SEO professionals because they need to be careful that the tactics they employ will not be considered spam by the search engines.

The Market Share of Search Engines

Figure 1-1 shows the U.S. market share for search engines throughout 2020, according to Statcounter. As you can see, Google is the dominant search engine on the Web worldwide.

Market share

100%

75%

50%

25%

0%

Jan '10 Jul '10 Jan '11 Jul '11 Jan '12 Jul '12 Jan '13 Jul '13 Jan '14 Jul '14 Jan '15 Jul '15 Jan '16 Jul '16 Jan '17 Jul '17

Oct '20

• Google   88.14%

• bing   6.18%

Yahoo!   2.51%

• Baidu   0.59%

• Yandex RU   0.26%

Google   bing   Yahoo!   Baidu   Yan

However, in some markets, Google is not dominant. In China, for instance, Baidu is the leading search engine. The fact remains, however, that in most world markets, a heavy focus on Google is a smart strategy for SEO.

Goals of Searching: The User's Perspective

The basic goal of a search engine user is to obtain information relevant to a specific set of search terms entered into a search box, also known as a query. A searcher may formulate the query as a question, but the vast majority of searches are performed by users simply entering word combinations—leaving the search engines to do the work of determining a query's "intent." One of the most important elements of building an SEO strategy for a website is developing a thorough understanding of the psychology of your target audience, and how they use words and concepts to obtain information about the services and/or products you provide. Once you understand how the average search engine user—and, more specifically, your target audience—utilizes query-based search engines, you can more effectively reach and keep those users.

Search engine usage has evolved over the years, but the primary principles of conducting a search remain largely unchanged. Most search engine use includes the following steps:

1. Users experience the need for information. They may be looking for information on a specific website, and they will search for that website (a navigational query); they might want to learn something (an informational query); or they might want to buy something (a transactional query). We will discuss this in more detail in the following section, Determing Searcher Intent.

2. Users formulate that need using a string of words and phrases (i.e. search terms), comprising the query. As shown in Figure 1.2, 58.8% of user search queries are one to three words long, though

as users are becoming more web savvy, they may use longer queries to generate more specific results more quickly.

3. Users execute the query, check the results, and if they seek additional information, will try a refined query.

Word Count	usa		gbr		deu		ind		can		fra	
	Raw	% of Total	Raw	% of Total	Raw	% of Total	Raw	% of Total	Raw	% of Total	Raw	% of Total
1	5,109,341	5.1%	2,129,285	5.4%	3,257,316	10.5%	1,885,665	7.0%	1,099,250	6.6%	992,680	8.7%
2	22,621,749	22.7%	9,614,489	24.6%	11,800,497	38.2%	7,381,115	27.4%	4,378,545	26.1%	3,481,288	30.4%
3	30,983,691	31.0%	12,412,277	31.7%	9,098,286	29.5%	7,939,094	29.4%	5,161,029	30.8%	3,412,031	29.8%
4	19,990,734	20.0%	7,791,976	19.9%	4,078,018	13.2%	5,014,621	18.6%	3,164,160	18.9%	1,970,486	17.2%
5	10,623,601	10.6%	3,821,129	9.8%	1,599,011	5.2%	2,536,011	9.4%	1,559,548	9.3%	922,326	8.1%
6	5,159,430	5.2%	1,732,546	4.4%	617,856	2.0%	1,168,400	4.3%	715,989	4.3%	380,772	3.3%
7	2,566,431	2.6%	822,879	2.1%	244,350	0.8%	546,821	2.0%	346,691	2.1%	153,310	1.3%
8	1,321,437	1.3%	406,466	1.0%	103,650	0.3%	245,967	0.9%	169,328	1.0%	65,103	0.6%
9	707,300	0.7%	209,727	0.5%	43,933	0.1%	115,020	0.4%	84,370	0.5%	31,584	0.3%
10	353,432	0.4%	95,429	0.2%	18,942	0.1%	56,558	0.2%	39,336	0.2%	15,111	0.1%
>10	426,584	0.4%	78,844	0.2%	16,666	0.1%	83,549	0.3%	33,472	0.2%	15,427	0.1%
Total	99,863,730		39,115,047		30,878,525		26,972,821		16,751,718		11,440,118	

Figure 1-2. Search query lengths

When this process results in the satisfactory completion of a task, a positive experience is created for the user, the search engine, and the site providing the information or result.

Determining User Intent: A Challenge for Search Marketers and Search Engines

Good marketers are empathetic, and smart SEO practitioners—as well as the search engines—share the goal of providing searchers with results that are relevant to their queries. Therefore, a crucial element to building an online marketing strategy around SEO and organic (also referred to as "non-paid") search visibility is understanding your audience and how they think about, discuss, and search for your service, product, and brand.

Search engine marketers need to be aware that search engines are *tools*—resources driven by intent toward a content destination. Using the search box is fundamentally different from entering a URL into the browser's address bar, clicking on a bookmark, or clicking on a link to go to a website. Searches are performed with *intent*—the user wants to find specific information, rather than just land on it by happenstance; search is also different from *browsing*, or clicking around links on a web page.

What follows is an examination of the different types of search queries and their categories, characteristics, and processes.

Navigational Queries

Users perform navigational searches with the intent of going directly to a specific website. In some cases, the user may not know the exact URL, and the search engine serves as the "White Pages." Figure 1-3 shows an example of a navigational query.

Figure 1-3. Navigational query

tunities
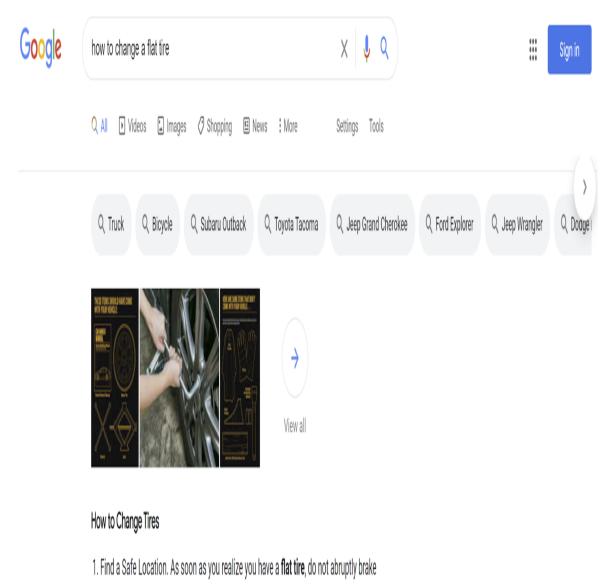
> Pull searcher away from destination; get ancillary or investigatory traffic. However, a September 2019 study from Perficient shows that nearly 70% of all clicks go to the first search result for branded queries. 1

Average traffic value:

> Very high when searches are for the publisher's own brand. These types of searches tend to lead to very high conversion rates. However, these searchers are already aware of the company brand, and some percentage of these queries may not represent new customers and for all of these queries the user began with an intent to visit the brand site. For brands other than the one being searched on, the click-through rates will tend to be low, but this may represent an opportunity to take a customer away from a competitor.

Informational Queries

Informational searches involve an incredibly broad range of queries. Consider the many types of information people might look for: local weather, driving directions, a celebrity's recent interview, disease symptoms, self-help information, how to train for a specific type of career...the possibilities are as endless as the human capacity for thought. Informational searches are primarily non-transaction-oriented (although they can include researching information about a product or service); the information itself is the goal, and in many cases no interaction beyond clicking and reading is required for the searcher's query to be satisfied. Figure 1-4 shows an example of an informational query.

🔍 Truck   🔍 Bicycle   🔍 Subaru Outback   🔍 Toyota Tacoma   🔍 Jeep Grand Cherokee   🔍 Ford Explorer   🔍 Jeep Wrangler   🔍 Dodge

→

View all

How to Change Tires

1. Find a Safe Location. As soon as you realize you have a **flat tire**, do not abruptly brake or turn. ...

2. Turn on Your Hazard Lights. ...

3. Apply the Parking Brake. ...

4. Apply Wheel Wedges. ...

5. Remove the Hubcap or Wheel Cover. ...

6. Loosen the lug nuts. ...

7. Place the Jack Under the Vehicle. ...

8. Raise the Vehicle With the Jack.

Figure 1-4. Informational query

Opportunities

Brand searchers with positive impressions of your site, information, company, and so on; attract inbound links; receive attention from journalists/researchers; potentially convert to sign up or purchase.

Average traffic value:
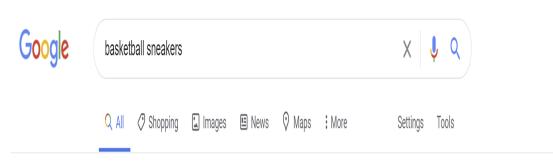
The searcher may not be ready to make a purchase, or may not even have long-term purchase intent, so the value tends to be "medium" at best. However many of these searchers will later perform a more refined search using more specific search temrs, which represents an opportunity to capture mindshare with those potential customers. For example, informational queries that are focused on researching commercial products or services can have high value.

Perficient, September 13, 2019, https://www.perficient.com/insights/research-hub/impact-of-search-features-on-ctr.
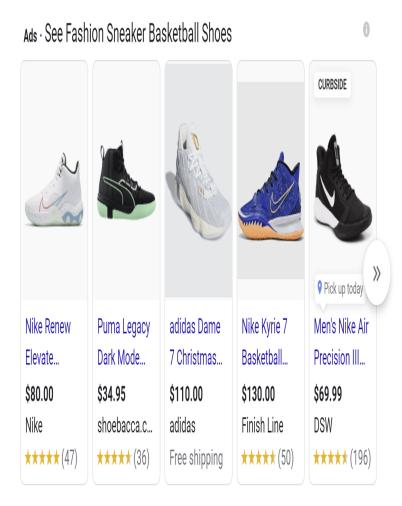
Transactional Queries

Transactional searches don't necessarily have to involve a credit card or immediate financial transaction. Creating a Pinterest account, signing up for a free trial account at DomainTools (http://www.domaintools.com), or finding the best local Japanese restaurant for dinner tonight are all transactional queries. Figure 1-5 shows an example of a transactional query.

Google

basketball sneakers   ✕ 🎤 🔍

🔍 All   🏷 Shopping   🖼 Images   📰 News   📍 Maps   ⋮ More   Settings   Tools

About 173,000,000 results (0.95 seconds)

www.nike.com › Basketball › Shoes ▾

Men's Basketball Shoes. Nike.com

Put in work and play hard with men's **basketball shoes** and sneakers from Nike.com.

Jordan · Low Top · Basketball · Mens Blue Basketball Shoes(23)

Opportunities

Achieve transaction (financial or other).

Average traffic value:
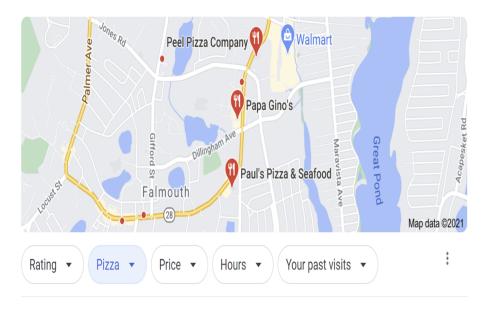
Very high. Transactions from these queries may not be immediate and it's up to the site receiving the related traffic to provide enough value to the user to convert them on their site or to make enough of an impression so that the user comes back and converts later.

Local Queries

As the name implies, local searches relate to users seeking information on things near to a specific location, such as where they currently are, or close to a location that they have referenced in the query. Examples might include looking for directions to the local park, a place to buy a slice of pizza, or the closest movie theater. Local queries are not an intent in the same way that navigational, informational, and transactional are, but represent a sub class that cuts across all types of queries. Many local queries are transactional but they differ as they relate to actions or transactions that will occur in person, but you can have navigational or informational local queries as well. Figure 1-6 shows an example of a local query.

Google   pizza   ✕ 🎤 🔍

🔍 All   📍 Maps   🖼 Images   📰 News   🏷 Shopping   ⋮ More   Settings   Tools

About 1,200,000,000 results (1.47 seconds)

Rating ▾   Pizza ▾   Price ▾   Hours ▾   Your past visits ▾   ⋮

Paul's Pizza & Seafood
4.5 ★★★★½ (256) · $$ · Pizza
14 Benham Rd
Closed · Opens 4PM
✓ Dine-in · ✓ Takeout · ✗ Delivery

Papa Gino's
4.0 ★★★★☆ (250) · $ · Pizza
56 Davis Straits
✗ Dine-in · ✓ Takeout · ✓ No-contact delivery

Peel Pizza Company
4.0 ★★★★☆ (26) · Pizza
31 Teaticket Hwy
Curbside pickup · Delivery · Takeout

Figure 1-6. Local query

Opportunities

> Drive foot traffic based on the proximity of the searcher. Offers a strong potential to achieve a transaction (financial or other).

Average traffic value:

> Very high. When users search on something near them the probability that they are interested in direct interaction, and possibly a near-term transaction is high. We can see that in the way that Google has tailored their SERPs for local queries to meet this demand.

Adaptive Search

The search engines also look at sequences of search queries to determine intent. You can verify this by trying search sequences where you search for related businesses one after another.

For example, if you're searching for a ball bearing manufacturer, you might try a search on "Schaeffler" and then follow that with a search on Timken (which is a ball bearings manufacturer). Normally these would simply behave as navigational searches and provide you with links to the company web sites. But, when you perform these two queries in succession, Google learns that you're in research mode and they provide you a modified form of result as seen in Figure 1-7, with many different ball bearings manufacturers shown in a carousel format at the top of the search results.

timken

Related to Timken Company and Schaeffler Group

TIMKEN	NTN	MOTION & CONTROL NSK	SKF	Koyo	NACHi	IKO	SCHAEFFLER
Timken Company	NTN Corporation	NSK	SKF	Koyo	Nachi-Fujikoshi	IKO	Schaeffler Group

Feedback

www.timken.com

The Timken Company | Engineered Bearings & Power ... ✓

Timken engineers and manufactures bearings and mechanical power transmission products. We use our knowledge to help global industries operate more ...

Search timken.com

Timken Careers ✓

Apply online for jobs at Timken -

Engineered Bearings ✓

Ball Bearings - Housed Units -

Timken
Company ≤

TIMKEN

Manufacturing company

🌐 timken.com ✓

Figure 1-7. Example of Adaptive Search

Searcher Intent

When you are building keyword research charts for clients or on your own sites, it can be incredibly valuable to determine the intent of each of your primary keywords. Table 1-2 shows some examples.

Term	Queries	Intent	Monetary value per visitor
Beijing Airport	5400	Navigational	Low
Hotels in Xi'an	110	Informational	Mid
7-Day China tour package	30	Transactional	High
Sichuan jellyfish recipe	53	Infoformational	Low

This type of analysis can help you determine where to concentrate content and links, as well as where to place ads.
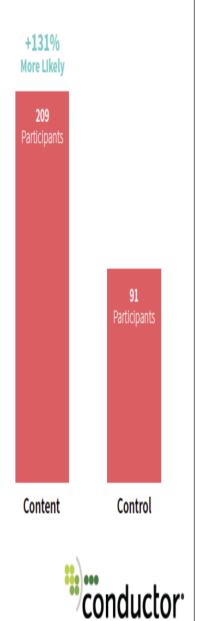
Hopefully, this data can help you to think carefully about how to serve different kinds of searchers based on their individual intents, and how to concentrate your efforts in the best possible areas.

Although informational queries are less likely to immediately convert into sales, this does not mean you should forgo pursuing rankings on these queries; getting your informative content in front of users seeking information can be incredibly valuable and can turn users into potential customers. As you'll see in Figure 1-8, data from a July 2017 study by Conductor showed that users who find useful informational content on your site are more likely to come to you to make a related purchase at a later date. They may also decide to share your information with others via their own website, or through social media engagement—an indirect but potentially more valuable result than converting the single user into a paying customer.

CONTENT VS. CONTROL: IMMEDIATELY AFTER READING A PIECE OF CONTENT

Immediately after reading a piece of content by the brand, consumers were **131% more likely to buy from that brand** compared to consumers who did not read any content.

Consumers were 131% more likely to buy after reading a piece of educational content

+131%
More Likely

209
Participants

91
Participants

Content

Control

conductor

Figure 1-8. How Informational Content Impacts User Trust

One problem in search is that when most searchers formulate their search queries, their input is limited to just a handful of words (Per Figure 1.2 78.8% of queries consist of one to four words). Because most people don't have a keen understanding of how search engines work, they often provide queries that are too general or that are presented in a way that does not provide the search engine (or the marketer) with what it needs to determine, with 100% accuracy 100% of the time, their specific intent.

Some search engine users may not have a specific intent behind a query beyond curiosity about a currently trending topic or a general subject matter. While this can make it challenging for a search engine to deliver relevant results, it poses a great opportunity for the digital marketer to capture the mind of someone who may not know what they are looking for - but who is interested in the subsequent variety of results the search engine delivers in response.

These types of general queries are important to most businesses because they often get the brand and site on the searcher's radar, which initiates the process of building trust with the user. Over time the user will move on to more specific searches that are more transactional or navigational in nature.

If, for instance, companies buying pay-per-click (PPC) search ads bought only the high-converting navigational and transactional terms and left the informational ones to competitors, they would lose market share to those competitors. Over the course of several days, a searcher may start with *digital cameras*, hone in on *Olympus OMD*, and then ultimately buy from the store that showed up in her search for *digital cameras* and pointed her in the direction of the Olympus OMD model.

To illustrate further, consider the case of user searches on the phrase "Ford Focus." They likely have numerous considerations on their mind when searching, even though they only use those two words in the query. Figure 1-9 illustrates what the rage of those considerations might be.
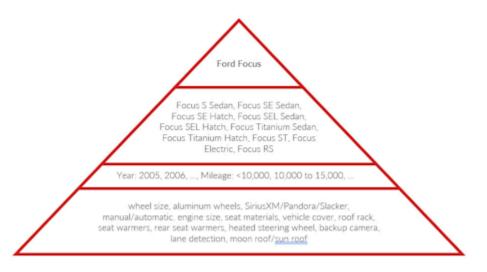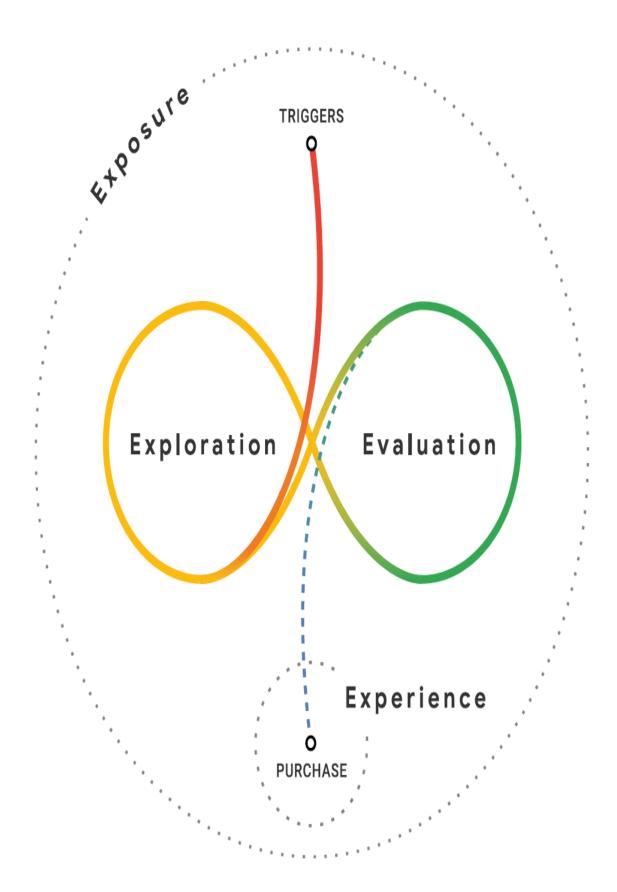
Figure 1-9. Pyramid of user needs

As we can see, the needs have many layers. The user who searched for "Ford Focus" may want a hatchback, a sedan, an electric car, or one of many specific model numbers. If they're buying a used car they may want to specify the year, or approximate mileage on the car. The user may also care about having aluminum wheels, Pandora, a roof rack, front & rear seat warmers, and various other options.

Recent research published by Think With Google generalizes this concept with something they refer to as the Messy Middle. As Figure 1-10 shows, this is the gap between the Trigger which causes the user to take action and the actual Purchase.

Exposure

TRIGGERS

Exploration

Evaluation

Experience

PURCHASE

Figure 1-10. Typical user journey to a purchase
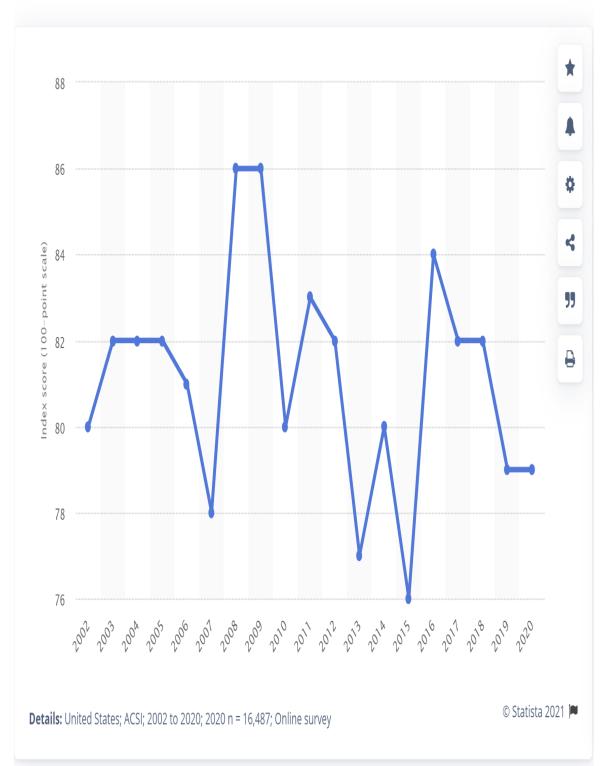
The Exploration/Evaluation part of this journey is highly complex and differs for every user. Whatever desires/needs users bring to this process, it's incumbent on the web site to try its level best to meet those desires and needs in order to earn the conversion.

Given the general nature of how query sessions start, though, determining intent is extremely difficult, and it can result in searches being performed where the user does not find what they want—even after multiple tries. Research from the American Customer Satisfaction Index (ACSI) found that 79% of Google users and 71% of Bing users were satisfied with their experiences. Figure 1-11 shows the ACSI satisfaction scores for Google from 2002 through 2020.

U.S. customer satisfaction with Google from 2002 to 2020

(index score)

Details: United States; ACSI; 2002 to 2020; 2020 n = 16,487; Online survey

Figure 1-11. User satisfaction with Google over time

While 79% satisfaction is an amazing accomplishment given the complexity of building a search engine, this study still showed that more than 20% of users were not satisfied with Google's search results; these numbers could reflect users' dissatisfaction with the number of ads that increasingly inflitrate the *search engine results pages* (SERPs).

The important takeaway here is that in all instances, determining searcher intent remains a challenge; and when the searcher's intent is broad, there is ample opportunity to leverage your content assets with SEO.

As an SEO practitioner, you should be aware that some of the visitors that you attract to your site may have arrived for the wrong reasons (i.e., they were really looking for something else), and these visitors are not likely to help you achieve your digital marketing goals. Part of your task in performing SEO is to maintain a high level of relevance in the content placed on the pages you manage, to help minimize this level of waste—while still attempting to maximize your overall presence in the SERPs and gain brand exposure.

How Users Search

Search engines invest significant resources into understanding how people use search, enabling them to produce better (i.e., faster, fresher, and more relevant) search engine results. For website publishers, the information regarding how people use search can be used to help improve the usability of a site as well as search engine compatibility.

User interactions with search engines can also be multistep processes, as indicated in the user search session documented by Microsoft and shown in Figure 1-12.

Inside a real query "session"

Example decision: Which shoes to buy?

Total task time: 55 minutes and 44 seconds

■ Dwell time

■ 21 sec | Merrell shoes | Search

■ 2 min **www.onlinestores.com**

■ 1 min **www.merrell.com**

■ 6 sec | Discount Merrell Shoes | Search

■ 4 min **www.nextag.com**

■ 12 sec | Merrell women's sandals | Search

■ 8 min **www.coachlikeapro.com**

■ 3 min | Clarks shoes | Search

■ 9 min **www.clarks.com**

■ 5 sec | Easy spirit | Search

■ 1 min **www.zappos.com**

www.easyspirit.com 27 min ▬▬▬▬ $

Figure 1-12. Merrell shoes user search session
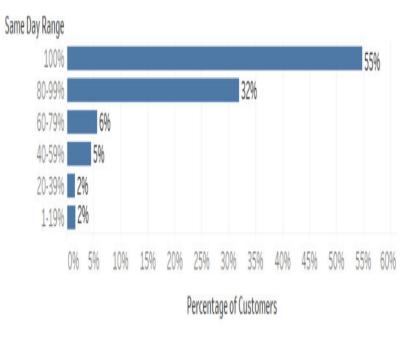
In this sequence, the user performs five searches over a 55+ minute period before making a final selection. The user is clearly trying to solve a problem and works at it in a persistent fashion until the task is done.

Ad Management platform provider Marin Software provided us with data that shows consumer latency in completing a purchase based on a review of all their clients. As you will see in Figure 1-13, for 55% of the companies in their data, the conversion happens the same day on which the initial visit occurs for 100% of customers. In other words, there is no latency at all for those companies.
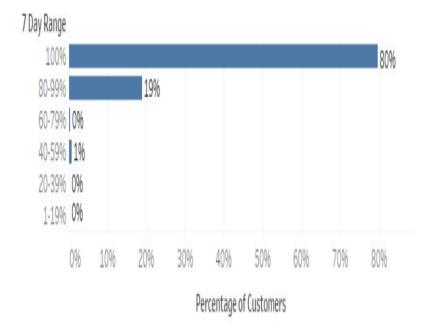
Percentage of Conversions Occuring on the Same Day as the Click

Same Day Range

Range	Percentage
100%	55%
80-99%	32%
60-79%	6%
40-59%	5%
20-39%	2%
1-19%	2%

Percentage of Customers

Percentage of Conversions Occuring within 7 Days of the Click

7 Day Range

Range	Percentage
100%	80%
80-99%	19%
60-79%	0%
40-59%	1%
20-39%	0%
1-19%	0%

Percentage of Customers

Figure 1-13. Latency in completing purchases

However, for 45% of the companies there is some latency in completion of purchases. This behavior pattern indicates that people are thinking about their tasks in stages. As in our Merrell shoes example in Figure 1-12, people frequently begin with a general term and gradually get more specific as they get closer to their goal. They may also try different flavors of general terms. In Figure 1-12, it looks like the user did not find what she wanted when she searched on *Merrell shoes,* so she then tried *discount Merrell shoes.* You can then see her refine her search, until she finally settles on Easy Spirit as the type of shoe she wants.

This is just one example of a search sequence, and the variety is endless. Figure 1-14 shows another search session, once again provided courtesy of Microsoft.

In this search session, the user has a health concern. This particular user starts with a five-word search, which suggests that she may have some experience using search engines. At 3:01, her search on *headache pregnant 3rd trimester* leads her to Yahoo! Answers (http://answers.yahoo.com). After visiting this site, the user suddenly gets more specific in her search.

She begins to focus on gestational diabetes, perhaps because something she saw on Yahoo! Answers led her to believe she may have it. The session culminates in a search for *first signs of gestational diabetes*, which suggests that she has concluded that this is quite possibly the issue she is facing.

The session stops there. It may be that at this point the user feels she has learned what she can. Perhaps her next step is to go to her doctor with her concerns, prepared to ask a number of questions based on what she has learned.

Typical session: health

10 unique queries
7 partial re-queries and refinements
57 minutes

2:58   Google   nausea headache pregnant what's wrong

---

pregnancy.families.com

2:58   nausea headache pregnant 3rd trimester

■ 2 min   **answers.yahoo.com/question**

3:00   Define malaise

▌ 1 min   **www.chefjournal.org/content**

3:01   Nausea headache pregnant 3rd trimester   Headache pregnant 3rd trimester

---

■ 5 min   **answers.yahoo.com/question**

3:06   gestational diabetes   gestational diabetes personal story

---

▬ 16 min   **www.experienceproject.com**

3:22   pregnancy complications stories

■ 4 min   **www.diabeticmommy.com/articles.html**

3:26   gestational diabetes itchy skin

▬ 14 min   **www.steadyhealth.com/Diabetes**

3:40   first signs of gestational diabetes

---

15 min   ▬   **www.articlebase.com;**
www.man-health-fitness-solutions.com;
www.healthspy.net; www.diabeteslife.com/diabetes;
www.optimumdiabetics.com;
www.everydayhealth.com;
www.squidoo.com; mascan-files-anti.com/200092/scan

Figure 1-14. Health user search session

Let's look at another search session example, . This session begins with a navigational search, where the user simply wants to locate the travel website Orbitz.com (http://www.orbitz.com) . The user's stay there is quite short, and she progresses to a search on *Cancun all inclusive vacation packages*. Following that, she searches on a few specific resorts and finally settles on *cancun riviera maya hotels*, after which it appears she may have booked her hotel—the final site visited on that search is for Occidental Hotels & Resorts (http://bookings.occidentalhotels.com), and the direction of her searches changes after that.

Typical session: travel

11 unique queries
5 partial re-queries and refinements
33 minutes

19:14 Google	Orbitz travel	
	▪ 1 min	**www.orbitz.com**
19:15 Google	Cancun all inclusive vacation packages	
	▬ 14 min	**cancun.bookit.com; www.travelocity.com; from bookit.com**
19:29 Google	Allegro Playacar - All inclusive	
	▪ 2 min	**from bookit.com; www.occidentalhotels.com**
19:31 Google	Occidental Grand Xcaret - All inclusive resort	
	▪ 3 min	**from bookit.com; search.bookit.com; www.occidentalhotels.com; from bookit.com**
19:34 Google	NH Riviera Cancun - All inclusive	
	▪ 1 min	**www.cancun.com/Hotels/NH-Riviera-Cancun; www.tripadvisor.com; www.travel-center.com**
19:35 Google	cancun riviera maya hotels	
	▬ 7 min	**www.nh-hotels.com; www.cancunandrivieramaya.com; www.nh-hotels.com; www.occidentalhotels.com; bookings.occidentalhotels.com**
19:42 YAHOO!	Cancun theme park Cancun theme park	
	▪ 1 min	**www.cancun.com/Ecological_Parks/Xcaret**
19:43 Live	xcaret	
	▪ 1 min	**search.live.com/results; search.live.com/video**
19:44 Live	xcaret	
	▪ 2 min	**www.xcaret.com; www.xcaret.com/Gallery.aspx; bookings.occidentalhotels.com/travel/itinerary.cfm; www.xcaret.com/Tickets/Xcaret_Basic_Entrance.html**
19:46 Live	xcaret video	
	▪ 1 min	**search.live.com/video; www.orbitz.com/App/View; www.orbitz.com/App/ViewPackageDetails**

Figure 1-15. Travel user search session

At that point, the user begins to look for things to do while she is in Cancun. She conducts a search for *cancun theme park* and then begins to look for information on *xcaret*, a well-known eco park in the area.

Users traverse countless different scenarios when they are searching for something. These example search sessions (Figures 1-14 and 1-15) represent traditional desktop interactions.
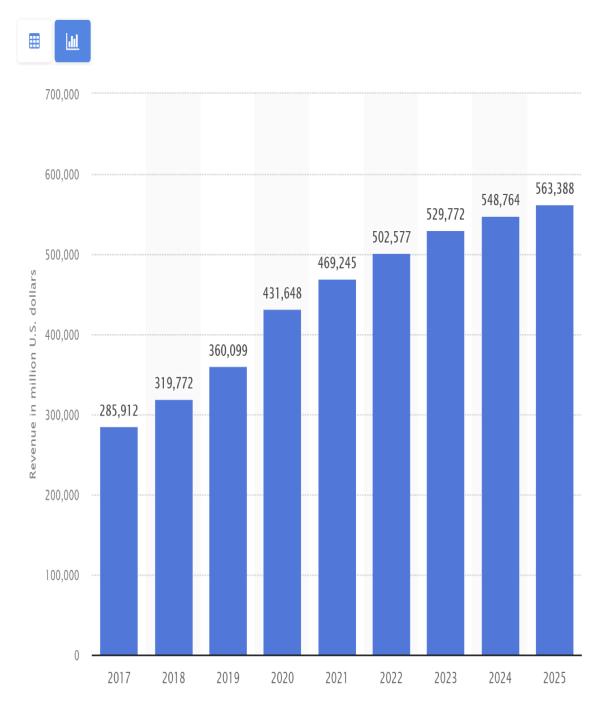
With respect to local search, data from Google states: "76% of people who conduct a local search on their smartphone visit a physical place within 24 hours and 28% of those searches result in a purchase." Search engines do a lot of modeling of these different types of scenarios to enable them to provide better results to users. The SEO practitioner can benefit from a basic understanding of searcher behavior as well. We will discuss searcher behavior in more detail in Chapter 2.

How Search Engines Drive Ecommerce

People make use of search engines for a wide variety of purposes, with some of the most popular being to research, locate, and buy products. US ecommerce sales reported by Digital Commerce 360 were $861 billion in 2020.[2] Statista forecasts that worldwide ecommerce retail sales will reach $6.5 trillion by 2023, as shown in Figure 1-16.

Retail e-commerce revenue in the United States from 2017 to 2025

(in million U.S. dollars)

Year	Revenue in million U.S. dollars
2017	285,912
2018	319,772
2019	360,099
2020	431,648
2021	469,245
2022	502,577
2023	529,772
2024	548,764
2025	563,388

It is important to note that search and offline behavior have a heavy degree of interaction, with search playing a growing role in driving offline sales. Figure 1-17 shows data from a May 2019 study by SEMrush which found that 27% of the traffic across their client base came from paid search.

Visualizing Search's Dominant Channel Share
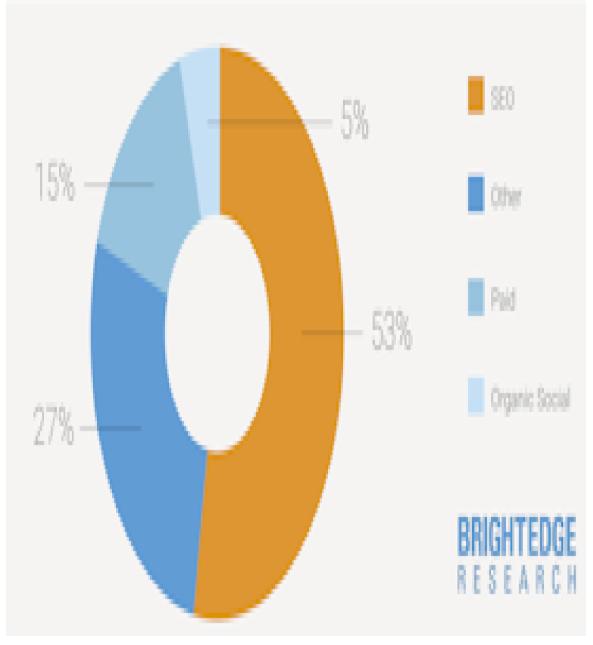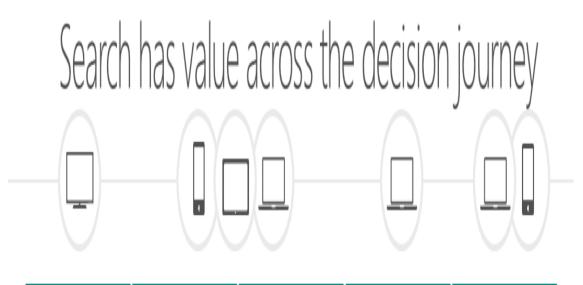
Figure 1-17. Sources of traffic to Brightedge customer sites

Driving traffic to your e-commerce site isn't just about driving conversions of every visitor. As shown in Figure 1-18, visitors that come to your site

from a search engine may be at any stage of the customer journey. This is why e-commerce sites should consider creating content for each and every stage. Even capturing those visitors when they are in the initial stages of discovery and research can significantly increase your chances of making them the sale when they are ready to buy.

Search has value across the decision journey

INITATION	RESEARCH	COMPARISON	TRANSACTION	EXPERIENCE
Goal: Get background information and buying landscape to become a more informed researcher	**Goal**: Explore buying guides, recommendations, and products that meet basic criteria	**Goal**: Deep comparison of a handful products that meet my criteria, including ratings, reviews, features, cost	**Goal**: Find where to buy, see pricing and promotions, availability, local stores	**Goal**: Get customer service, ask maintenance questions, make additional purchases

Type of Searches:

BRAND SEARCHES

CATEGORY SEARCHES

TANGENTIAL SEARCHES

COMPETITOR BRAND SEARCHES

Click Tracking: How Users Click on Results (Organic Versus Paid)

By now, you should be convinced that you want your site content to be prominently displayed within SERPs.

In contrast, data shows that you may *not* want to be #1 in the paid search results, because the resulting cost to gain the #1 position in a PPC campaign can reduce the total net margin on your campaign. As shown in Figure 1-19 London based Hallam Internet published data in 2019 that showed that the 3rd and 4th ad positions may offer the highest ROI:
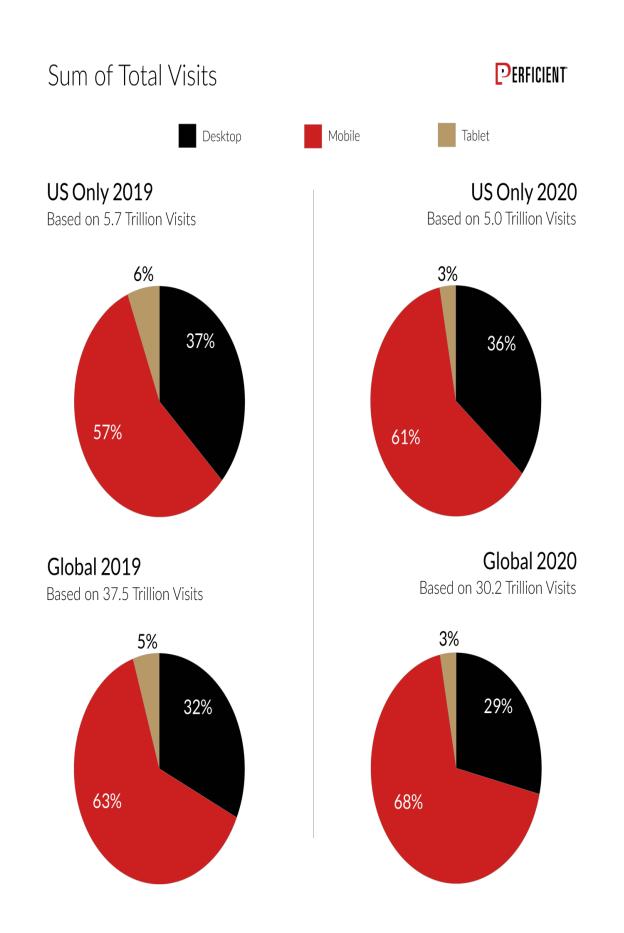
Figure 1-19. Organic search delivers traffic at all stages of the customer journey

Please source this to: https://www.hallaminternet.com/google-adwords-which-ad-position-is-best/
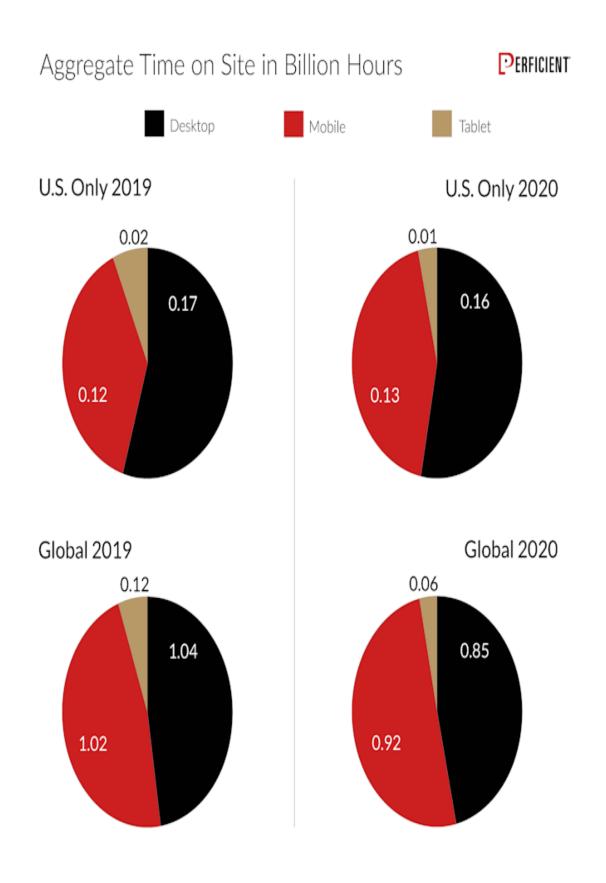
Of course, many advertisers may seek the #1 position in paid search results, as benefits include branding, and maximizing market share. For example, if an advertiser has a really solid backend on their website and are able to make money when they are in the #1 position, they may well choose to pursue it. Nonetheless, the data from the Hallam research suggests that due to the lower ROI there are many organizations for which being #1 in paid search does not make sense.

Search Traffic by Device

The majority of traffic to web sites today comes from mobile devices. In 2021 Perficient conducted a comprehensive study of mobile vs. desktop traffic (https://www.perficient.com/insights/research-hub/mobile-vs-desktop-usage-study) which showed that in the US 61% of all traffic comes from mobile devices, and globally that number is 68%, with tablets accounting for another 3% of overall traffic in both views of the data (as shown in Figure 1-20).

Sum of Total Visits

P ERFICIENT

■ Desktop   ■ Mobile   ■ Tablet

US Only 2019
Based on 5.7 Trillion Visits

6%
37%
57%

US Only 2020
Based on 5.0 Trillion Visits

3%
36%
61%

Global 2019
Based on 37.5 Trillion Visits

5%
32%
63%

Global 2020
Based on 30.2 Trillion Visits

3%
29%
68%

Figure 1-20. Mobile vs. Desktop Traffic, US and Global Views

This does not mean that desktop has become unimportant, as it delivers the most total time on site from visitors (in aggregate) and nearly the same number of page views as mobile devices. Figure 1-21, also from the Perficient study, shows the aggregated (across all visitors) total time on site for desktop vs. mobile.

Aggregate Time on Site in Billion Hours

PERFICIENT

■ Desktop　　■ Mobile　　■ Tablet

U.S. Only 2019

0.02

0.17

0.12

U.S. Only 2020

0.01

0.16

0.13

Global 2019

0.12

1.04

1.02

Global 2020

0.06

0.85

0.92

Figure 1-21. Mobile vs. Desktop aggregate time on site, US and Global Views

Google has been long aware of the growing importance of mobile devices and first announced an algorithm update to focus on mobile devices in April 2015. This announcement was dubbed "Mobilegeddon" by the industry as many expected the impact of the update to be a dramatic upheaval of the search landscape. In reality it's almost never going to be in Google's interest to completely disrupt the existing search results in a dramatic fashion because as we discussed earlier in this chapter user satisfaction with Google has stayed at roughly the 80% level year over year since 2002.

This is reflected in how Google handled its switch to Mobile-first Indexing. This was first announced by Google on November 4, 2016 (https://developers.google.com/search/blog/2016/11/mobile-first-indexing). The initial announcement stated:

Today, most people are searching on Google using a mobile device. However, our ranking systems still typically look at the desktop version of a page's content to evaluate its relevance to the user. This can cause issues when the mobile page has less content than the desktop page because our algorithms are not evaluating the actual page that is seen by a mobile searcher.

Nearly everyone in the industry expected this to unfold far more quickly than it did. The reason that it didn't is that Google has to perform extensive amounts of testing on any change it makes to how it indexes and ranks content in order to minimize unintended negative impacts on their search results. Due to the scale of search this is a highly involved and cumbersome process.
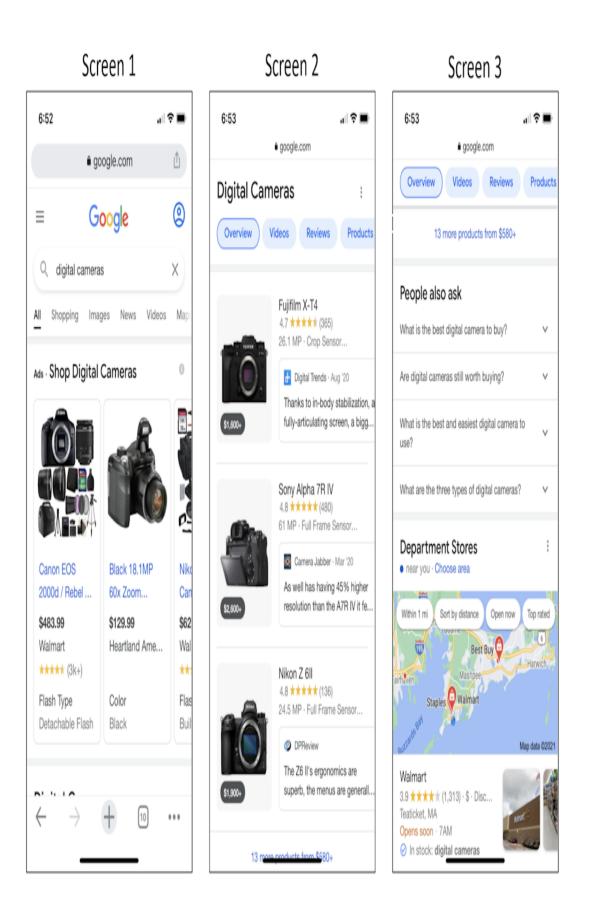
It was not until March 5, 2020 that Google announced a target date to make the switch 100% universal with a switch date of September 2020 – nearly four years after the initial announcement (https://developers.google.com/search/blog/2020/03/announcing-mobile-first-indexing-for).This announcement stated:

It's been a few years now that Google started working on mobile-first indexing - Google's crawling of the web using a smartphone Googlebot. From our analysis, most sites shown in search results are good to go for mobile-first indexing, and 70% of those shown in our search results have already shifted over. To simplify, we'll be switching to mobile-first indexing for all websites starting September 2020. In the meantime, we'll continue moving sites to mobile-first indexing when our systems recognize that they're ready.

From an SEO perspective this means that Google is going to crawl the mobile version of your site and analyze the structure and content of your mobile site to determine the types of queries for which your site is relevant. This means that the great majority of your SEO focus needs to be on the mobile version of your site.

The fact that most visitors to your site come from mobile devices is of critical importance. Among other things, it means that website design should start with mobile functionality, design, and layout . Any other approach is likely to result in a mobile site that is not as optimal as it could be. It also means that for most queries that you should be studying the structure and format of the mobile search results from Google. Figure 1-22 shows us sample search results for the query "digital cameras".

Screen 1

6:52

🔒 google.com

☰   **Google**   ⊙

🔍 digital cameras   ✕

All   Shopping   Images   News   Videos   Map

Ads · **Shop Digital Cameras**   ⓘ

Canon EOS 2000d / Rebel ...	Black 18.1MP 60x Zoom...	Niko... Cam...
$483.99	$129.99	$62
Walmart	Heartland Ame...	Wal...
★★★★★ (3k+)		★★
Flash Type	Color	Flas...
Detachable Flash	Black	Buil...

Digital C

← → ＋ 🔟 •••

Screen 2

6:53

🔒 google.com

Digital Cameras   ⋮

Overview   Videos   Reviews   Products

Fujifilm X-T4
4.7 ★★★★★ (365)
26.1 MP · Crop Sensor...

📘 Digital Trends · Aug '20

Thanks to in-body stabilization, a
fully-articulating screen, a bigg...

$1,800+

Sony Alpha 7R IV
4.8 ★★★★★ (480)
61 MP · Full Frame Sensor...

📷 Camera Jabber · Mar '20

As well has having 45% higher
resolution than the A7R IV it fe...

$2,600+

Nikon Z 6II
4.8 ★★★★★ (136)
24.5 MP · Full Frame Sensor...

🅳 DPReview

The Z6 II's ergonomics are
superb, the menus are generall...

$1,900+

13 more products from $580+

Screen 3

6:53

🔒 google.com

Overview   Videos   Reviews   Products

13 more products from $580+

People also ask

What is the best digital camera to buy?   ⌄

Are digital cameras still worth buying?   ⌄

What is the best and easiest digital camera to
use?   ⌄

What are the three types of digital cameras?   ⌄

Department Stores   ⋮

● near you · Choose area

Within 1 mi   Sort by distance   Open now   Top rated

Map data ©2021

Walmart
3.9 ★★★★★ (1,313) · $ · Disc...
Teaticket, MA
Opens soon · 7AM
✓ In stock: digital cameras

Figure 1-22. Mobile SERPs for the query "digital cameras"

In the first three screens of results on mobile devices, the ranking opportunities are in Google Shopping, People Also Ask Boxes, and Local Search. A core part of your SEO strategy is to develop an understanding of the search landscape at this level of detail, as it can directly impact the search terms you're choosing to target.

Targeting informational queries is quite different, as shown in Figure 1-23. Here you still see Google Shopping at the top of the results, but the next two screens are filled with access to informational content. As a result, the nature of the ranking opportunities is quite different.

Screen 1

8:51

🔒 google.com

☰   **Google**   👤

🔍 history of prague   ✕

All   Videos   Images   News   Maps   Shopping

Ads · Shop history of prague   ⓘ

(Hardcover) (Audiobooks) (eBooks) (Paperback) (A

A History of the Czech Lands...

$31.19
Google Play
Free shipping

Format
eBook

Prague in Black and Gold: The...

$19.39
Amazon.com

Format
Paperback

Art ...
Prag...

$4.1
Thri
Use

For
Pap

← → ＋ 🔟 •••

Screen 2

8:51

🔒 google.com

History of Prague   ⋮

The history of Prague covers more than a thousand years, during which time the city grew from the Vyšehrad Castle to the capital of a modern European state, the Czech Republic. Wikipedia

Videos

YouTube · Pražský cyklista   YouTube · Discerni

20:14   PREVIEW   3:35

Czech documentary about the history and origin of Pra...

Introduction to Bohemia

← → ＋ 🔟 •••

Screen 3

8:52

🔒 google.com

People also ask   ⋮

What is Prague famous for?   ⌄

Was Prague ever part of Germany?   ⌄

How did Prague get its name?   ⌄

What was Prague before?   ⌄

Feedback

Ⓦ en.m.wikipedia.org › wiki › History...   ⋮

History of Prague - Wikipedia

In the 13th century, towns started to increase in size. Three settlements around the Prague Castle gained the privilege of a town. The settlement ...

(Prehistory) (Medieval Prague) (Renaissance) (Jew

www.prague.com › history   ⋮

Brief History of Prague, Czech Republic | Prague.com

The city bloomed in the 14th century under the

← → ＋ 🔟 •••

In 2019, Perficient published the most recent comprehensive study of Google CTR (click-through rate), which shows CTR data for both desktop and mobile searches, as well as for branded and unbranded searches.[3] Perficient found that the disparity in CTR for branded terms and non-branded terms is truly remarkable. Figure 1-24 shows the average CTR (percentage of clicks received) by Google SERP position for both branded and non-branded queries.
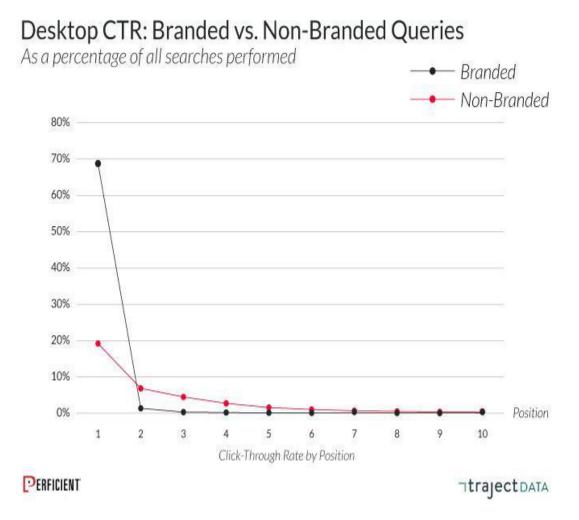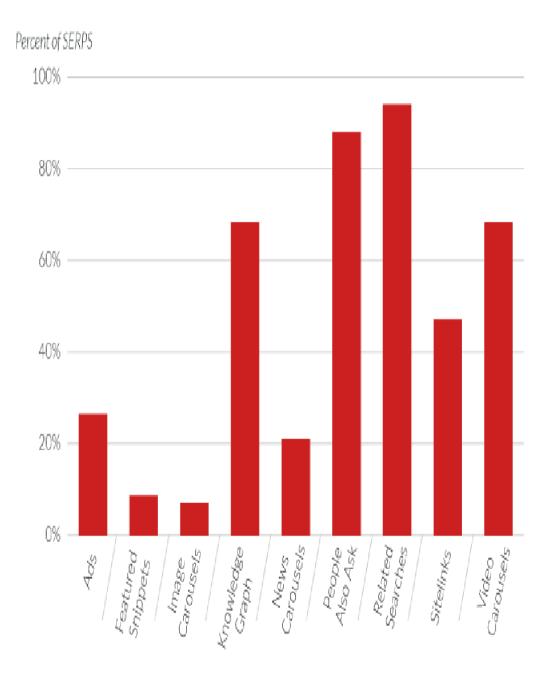

Figure 1-24. Search results CTR by ranking position

Today's Google search results are rich in search features of many different kinds. Figure 1-25 shows the frequency of different search features within

the Google SERPs as of December 2019.

Popularity of Search Features

Percent of SERPS

Figure 1-25. Popularity of search features

Each of these features creates different opportunities for placement in the search results and impacts the potential CTR you might experience. The reason that CTR is impacted is that users respond to different visual elements and their eyes get drawn to images and parts of the page that look different.

Way back in 2006, a research firm Enquiro (later acquired by Canada's Yellow Page Group) conducted heat-map testing with search engine users that produced fascinating results related to what users see and focus on when engaged in search activity. Figure 1-26 depicts a heat map showing a test performed on Google; the graphic indicates that users spent the most amount of time focusing their eyes in the upper-left area, where shading is the darkest. This has historically been referred to in search marketing as the "Golden Triangle."

IMAGE TO COME

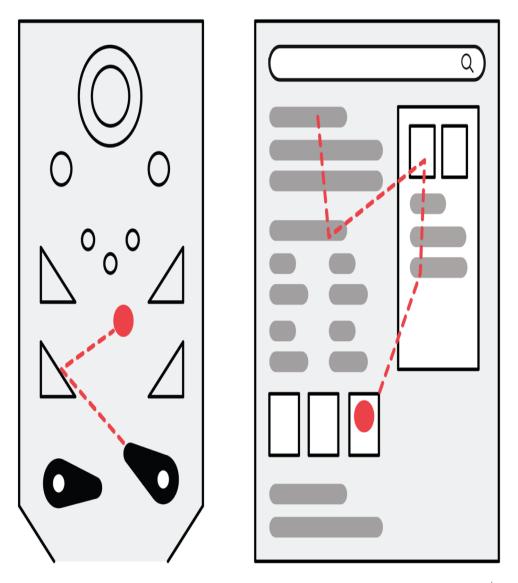*Figure 1-26. Enquiro eye-tracking results, 2006*

However, the search landscape has changed dramatically since 2006, and has become increasingly more complex - evolving towards results that are media rich, and mobile-centric.

As a result, the appearance of the search results is nowhere near as consistent as it used to be, which causes users to have a much less consistent approach to how their eyes scan the search results page. As shown in Figure 1-27, a more recent research written up by Search Engine Land and performed by Yext shows that users follow a pinball-like path through the SERP.

The Pinball Pattern

PINBALL
MACHINE

GAZE
PLOT

Figure 1-27. How users view search results

These types of studies illustrate how important search engine results pages really are. And, as the eye-tracking research demonstrates, as search results continue to evolve, users' search and engagement patterns will follow suit. There will be more items on the page for searchers to focus on, more ways for searchers to remember and access the search listings, and more interactive, location-based delivery methods and results layouts—which will continue to change as other search environments and platforms continue to evolve.

Role of AI/Machine Learning

By now you've got the idea that user behavior is highly complex and the challenge of meeting those needs with a search engine is enormous. As the seoClarity data in Figure 1-2 showed that 78.8% percent of all search queries consist of 4 words or less. These short phrases are all that a search engine gets in order to determine what results to return in the SERPs. In addition, March 2021 data from ahrefs shows that 92% of all search queries get ten searchers per month or fewer. That does not provide search engines with a lot of prior history in order to model what the user wants.

Google deploys a large array of resources to try and meet these challenges. This includes a growing number of machine learning algorithms. Some of the more well-known machine learning algorithms that Google has deployed are:

- RankBrain

- Bert

- SMITH

- MUM

You can expect to see Google continuing to roll out machine learning algorithms on an ongoing basis. This brings some unique challenges as the

nature of how these algorithms work is opaque to those that create them. As a result, testing them and validating that they work as expected is far more complex.

Further, these machine algorithms are being used to supplement the human-generated algorithms that Google developed over decades. Google will continue to work on improving these algorithms as well.

Conclusion

Search is an integral part of the fabric of global society. The way people learn, work, share, play, shop, research, socialize, and interact has changed forever, and organizations, causes, brands, charities, individuals—almost all entities—need to view their Internet presence as a core need, and they must leverage the search engines and search functionality within all platforms to generate exposure and facilitate engagement. This book will investigate further just how search, and therefore SEO, is at the center of the web ecosystem - and can play a major role in your success within the ever-evolving digital economy.

---

1 worldometer, January 10, 2021, https://www.worldometers.info/.   2 Brightedge, January 2020, https://www.brightedge.com/blog/organic-share-of-traffic-increases-to-53.   3 Statcounter, December 2020, https://gs.statcounter.com/search-engine-market-share.

2 Digital Commerce 360, "Quarterly Retail E-Commerce Sales, 1st Quarter 2015," November 19, 2020, https://www.digitalcommerce360.com/article/us-ecommerce-sales/.

3 Perficient, "How Organic Search CTR is Impacted by Google's Search Features," https://www.perficient.com/insights/research-hub/impact-of-search-features-on-ctr.

CHAPTER TWO

Search Fundamentals

Search has become a fundamental aspect of how we find information, identify solutions to problems, and accomplish various tasks ranging from making purchases to booking travel. Without search engines - whether standalone (e.g. Google) or platform specific (e.g. Twitter Search) – we would be unable to quickly find what we need, and to execute the tasks we seek to accomplish. We often don't know where to find an answer or solution, or whether one exists - and even when you know where something is located, it can still be more efficient to use a search engine to retrieve it.

For instance, consider this URL:

```
https://blogs.example.com/archive/articles/2019/dogs/canine-
dentistry-advice.xhtml
```

This web page may have information you're looking for pertaining to canine dentistry. However, you likely have no idea that this blog exists – let alone where to find this specific page within the blog, given the complexity of the URL structure. Additionally, given its age (2019) it's likely that this article would be 'archived,' which means it wouldn't necessarily be on the blog's home page – you'd have to drill down at least three levels, and you might have to sift through dozens or hundreds of articles in the final list.

Conversely, consider the keystrokes required to go to www.google.com and perform a search for *"canine dentistry advice."* Assuming that the above example page is in Google's index, and that Google considers it a relevant result for you when you perform the search, it would be much more efficient to search for this topic than to type in the exact URL into the browser address bar, or to navigate to it through the blog's post archives.

Thus, Google is not solely a provider of potential answers; it is also the interface to the Internet, and a method of navigating websites, many of which may be poorly designed. In essence, *search engines enable us to connect efficiently to information, people, and actionable online activities.* Not simply to find 'things' – but to connect people to their interests by way of sites, pages, and services.

The Internet, including the people and things which connect to it, is a system comprised of purposeful connection among disparate components and resources. As an SEO practitioner within this system, you can influence the interconnectedness of these various components to present quality, relevant, authoritative, trustworthy, and useful content, and you can make your website easy to navigate and understand, for both users and Google.

To be an SEO is to be a master of organization and connections – between searches and conversions, customers and products, web content and search engines, users, and influencers, and even between you and the people and organizations responsible for developing, maintaining, and promoting the site you're optimizing.

You also need to develop a comprehensive understanding of the topic matter of your website, what user needs relate to your market area, and how

users typically communicate those needs, which is expressed in the search queries they use. This becomes a guide to what, and how much, content you need to create to be successful in SEO.

For those of you seasoned in SEO, or for those who simply have high digital literacy, some of the material in this chapter may seem obvious or elementary at a glance – perhaps even arcane and obscure. However, context is essential in SEO, as understanding the digital ecosystem, including the people who connect to and through it, will help you understand how to present your site as a valuable, authoritative, and trustworthy resource on the Web for your topic area.

Deconstructing Search

While search technology has grown at a rapid pace, the fundamentals of search remain unchanged: a user performs a search (referred to as a "query") using a search engine, and the search engine generates a list of relevant results from its index and presents the results to the user on what is referred to as a search engine results page ("SERP").

We can divide the user search process into three fundamental components – the query, the index, and the results. Within each of these concepts there exist vast realms of information and knowledge that the SEO professional must possess. As a starting point, it is essential to begin to think of SEO in the following terms: search queries map to topics, which maps to keyword and topic research; a search engine's index maps to all aspects of technical SEO; and a given set of search results maps to content marketing and user behavior.

Let's start by analyzing the search query – the human interface to a search engine.

The Language of Search

Keywords are the common language between humans and search engines. On a basic level, a search engine seeks to understand the intent of the user's

search query, and then identify web pages (or other forms of web content) which best meet that intent. This is done by a variety of techniques including natural language processing (NLP).

In the old days of SEO, keywords were mostly individual words and combinations of words (phrases) minus **stop words** (short words like *to*, *of*, and *the*). In the modern era, keywords are often interpreted less literally by search engines as they seek to understand intent of the user in the context of understanding language meaning beyond exact word matches, while utilizing various additional information about that user to help inform the process of understanding user intent. Google's most advanced search system (as of 2022) named "MUM" is a multimodal training system comprised of various algorithms supposedly 1,000 times as powerful as its predecessor BERT[1]. MUM represents Google's evolution towards understanding information across languages and across content types while utilizing machine learning to enable their systems to learn as searches are performed, in order to better understand user intent, and to identify the most relevant information available on the web to return to that user within their search results.

As users, we tend to simplify our language when performing searches on search engines, in hopes that they can understand us better. For instance if you wanted to know how many Formula One world championships the Williams Racing team has won, your search query might look like this:

```
williams F1 "world championships"
```

While wouldn't use this language to ask a person about this, you have learned that search engines understand this language perfectly. As you will see in Figure 2-1 (ch2-one-box-result.jpg) Google provides a direct answer for that query (in the form of a **OneBox result**, which is defined later in this section) because it has high confidence in its interpretation of your intent.

Google

All | News | Images | Shopping | Videos | More | Tools

About 504,000 results (0.99 seconds)

Williams Grand Prix Engineering

2022 Formula One World Championship

Constructors' Championships	9 (1980, 1981, 1986, 1987, 1992, 1993, 1994, 1996, 1997)
Drivers' Championships	7 (1980, 1982, 1987, 1992, 1993, 1996, 1997)
Race victories	114
Podiums	313

21 more rows

https://en.wikipedia.org › wiki › Williams_Grand_Prix_E...

Williams Grand Prix Engineering - Wikipedia

About featured snippets · Feedback

Figure 2-1.

For simple queries, however, the search engines have evolved to the point that we can usually communicate with them in a more human-language way when our search intent is unambiguous and there is a definitive answer to our question. For instance, you would be served the same OneBox result for this query as you would for the natural language query: *how many F1 world championships does Williams have?*

In the first query, we translated our intent into keywords, then arranged them so that the search engine would interpret them properly. In the second query, we asked the question in natural, person-to-person language format. When conducting keyword research, you must consider both of these types of search behaviors. It is important to identify the words and phrases which exist to describe the topics related to your business – but it is just as important to understand the actual questions people ask the search engines when looking for the types of information and content you provide.

Word Order and Phrases

While different queries may return similar (or the same) results, word order is still an important factor in search queries and can influence the content, type, and ordering of the search results you receive. As an example, currently, while in the same search session you may be shown the same search results for *williams F1 "world championships"* as you would for *F1 williams "world championships"*, only the former query will display a OneBox result with the answer. Similarly, if you were to use the singular *"world championship"* instead of the plural, you may receive no OneBox answer.

Variations of keywords such as synonyms, misspellings, alternate spellings, and plurality are also considered by the search engines when determining searcher intent and rendering search results pages. However, you don't necessarily need to use every keyword variant you see in your keyword research within your content in order for the search engines to understand that your content is relevant to a query. For example, as shown in Figure 2-2 (ch2-resume-writing-services.jpg) if you offer resume writing services

you can find many closely related keyword variants. You don't need to use every single one of those variants on your page or have a page for each of these variants.

resume writing services	c	12,100
resume writing	c	8,100
resume writing service	c	8,100
writing a resume	I	3,600
best resume writing service	c	2,900
professional resume writing services	c	1,900
resume writing near me	T	1,900
professional resume writing	c	1,600
resume writing services near me	T	1,600

Figure 2-2.

Nonetheless, comprehensive keyword research is still important. When creating your initial keyword list, a good first step is to list all of your products, services, and brands, but the next step should be to consider other

words that would appear before and after them in a search query. This involves thinking about the larger topical context around your keywords, and their relationships to other topics.

To help with that profile all the user needs that relate to your products and/or services. How does the user think about what they want? What are the initial queries they might start with and how will they progress as they move through the purchase funnel? As you perform your keyword research keep all of these considerations in mind and make a plan to address as many of these user needs as possible.

Traditionally, search engines have ranked pages based on the similarity between the user's search query and the content they see on relevant web sites. If a search query appears verbatim in a page's content, that page has a stronger chance of being near the top of the results, but the use of related words and phrases is also important and further reinforces the context.

NOTE

Be wary of **over-optimizing** for a keyword. If it's obvious that a page has **thin content** that repeatedly and awkwardly uses keywords, Google will penalize the site by lowering its rankings or even removing its pages from the index. This topic is covered in more depth in Chapter 9.

Search Operators

When used correctly in a query, search operators are directives that limit the scope of the returned search results in specific ways, making them extremely useful for various forms of research and information retrieval.

Google doesn't provide a comprehensive list of search operators and the ones that work change from time to time, so the best general approach to identify available search operators is to use the **after:** search operator to limit the age of the pages in the result set:

```
"google search operators" after:2022-06-01
```

Here are a number of important search operators that have remained fairly stable in functionality over time:

OR: Includes specific alternatives for a given keyword. Can be expressed as **OR** or | (pipe symbol).

```
Christmas gifts for kids | children | boys | girls
```

NOT: The opposite of **OR**. Expressed as a – (dash) symbol before a keyword, with no space between:

```
Christmas gifts for kids -teens
```

site: Limits the search scope to the specified domain:

```
"the art of seo" site:oreilly.com
```

filetype: Limits the search scope to documents that have the specified three-letter file extension:

```
evil plan to destroy the linux operating system filetype:doc
site:microsoft.com
```

> **NOTE**
>
> As long as they aren't mutually exclusive, you can usually use more than one operator in a query.

Wildcard: When there are too many OR operators, you can use a wildcard (expressed as the ***** symbol) to indicate an unknown word within the context of a query:

```
evil plan to destroy the * operating system filetype:doc
site:microsoft.com
```

You can also use it to indicate all possible top-level domains or subdomains with the **site:** operator:

```
evil plan to destroy the linux operating system filetype:doc
site:*.microsoft.com
```

cache: Brings you directly to the most recent Google cache copy of the specified URL, if there is one:

```
cache:https://www.example.com/secrets/OMG_DELETE_THIS.htm
```

inanchor: Searches for a word used in anchor text (link text). You'd usually use this with the **site:** operator:

```
site:*.microsoft.com inanchor:monopolistic
```

allinanchor: Same as above, but for multiple words (notice that there's a space after the colon this time):

```
site:*.linux.com allinanchor: steve ballmer monkey dance
```

We leave the rest of the list – whatever it may be at the moment you read this – up to you.

Vertical and Local Intent

> **NOTE**
>
> This is a simplified overview of local, mobile, and vertical search. For more comprehensive coverage, refer to Chapter 11.

The concept of 'searcher intent' is extremely important to search engines because if the search engine accurately identifies user intent, and therefore serves relevant search results which satisfy that intent, then users will return to the search service for future needs – which results in an increased user base and increased advertising revenue (e.g. via Google Ads, for Google). As a result, search engines have a powerful incentive to correctly identify and understand user intent.

When someone searches for a specific kind of content (images, videos, events, news, travel, products, music, etc.), this is referred to a a **vertical** search, even if it's not performed from a vertical search engine. (This is explained in more detail later in this section).

A query can suggest vertical intent when it isn't explicitly requested. For instance if you search for *diamond and emerald engagement ring*, Google may interpret this query to mean that you want to see high-quality photos along with jewelry retail product page links, so your search results may include image and product results. If you search for *parrot playing peekaboo*, Google may interpret this query to mean that you're looking for videos, and may return videos prominently in your search results. Importantly, Google will factor in any available user behavior data about the user performing the search, to enhance their understanding of your query's intent.

Signals of local search intent are usually in the form of keywords, but the query does not necessarily need to include the name of a specific city or town for a search engine to interpret the query as having local intent. For example, when a user searches for "pizza delivery" without including any location in the query, a search engine will likely interpret this query as having local intent and utilize any information it has about the user to identify where they are – e.g. the IP address or other location data gleaned from the device being used by the user performing the search. Depending on how much location data access you allow Google to have, a mobile search with obvious local signals (for instance, using the keywords *near me* in your query) will often generate results that are relevant to the immediate area around you.

Local scope can be as specific as a street corner, or as broad as any place that has a name (a state, country, or continent). For instance:

- café near the Louvre
- best golf course in Pennsylvania
- nearest Ferrari dealership

- Japanese food in Disney Springs

- most expensive place to live in America

Crawling

Web crawling is the process that search engines use to discover content on your site and across the web (generally located on specific URLs). As a core component of information and content discovery for search engines, crawling plays a critical role in how search engines build their indices of web documents (URLs). As a result, making your site easy for search engines to crawl (making your site "crawlable" or "crawler-friendly") is a critical area of focus for your content development to support your ongoing SEO efforts. The topics of search engine crawling and indexation are described in more detail in Chapter 6.

Overall, the web is too vast for any search engine or any other entity to crawl completely, so search engines like Google need to prioritize crawler efficiency and effectiveness by limiting how much content they crawl, and how often. As a result, there is no guarantee that a search engine crawler will crawl all of your site's content - especially if your site is quite large.

There are many reasons why search engine crawlers may not crawl areas of your site.

These include:

1. The crawler never finds a link to the URL, and it does not appear in your XML site map file(s). (Perhaps you actually do link to the page but it's only accessible via JavaScript that does not render on your page until a user clicks on a page element).

2. The crawler becomes aware of the URL, but it is far down in your hierarchy, e.g. the crawler has to crawl too many other pages to reach it, and decides not to crawl the URL.

3. The crawler has crawled the page at some point in the past, and based on the search engine's interpretation of the page's content or

other characteristic, decides that there is no need to crawl it again.

4. Your site may be assigned a limited crawl budget and there is not enough available for the crawler to reach all of your site's content. This can happen for various reasons – including, for example, if there are issues with the server hosting your site at the time of the crawl, if there are multiple URLs which contain the exact same content, and other factors.

Ensuring the crawlability of your URLs requires understanding the development platform and environment for your content, how URLs are constructed and implemented, how redirects are handled and maintained, and numerous other factors discussed further in Chapter 6.

The Index

Today 'the search index' generally refers to the index-of-indexes, which contains metadata on many more asset types than HTML pages, including images, videos, PDF documents, and other file types. When we say a site is 'indexed,' that means that a search engine has connected to it through some means (a link from a page already in the index, or a sitemap submitted through the search engine's backend channels), **crawled** it with a script that discovers all links to find new content, performed semantic analysis of its content to understand its relevance, and created some descriptive metadata about it so that it can be associated with the words and intentions of searchers.

In addition to relying on information about URLs in its search index, Google also uses various data it obtains and stores about users to determine the relevance of indexed content to a search query.

The Knowledge Graph

The **Google Knowledge Graph** is a rapidly-evolving graph database that Google uses to understand how topics and concepts relate to one another. The Knowledge Graph is composed of trusted facts and their relationships. It was originally populated by structured data from reliable public sources

such as Wikipedia, Wikidata, and *The CIA World Factbook*. Today Google also incorporates structured data from many other sites and services and uses machine learning to analyze and collect data from search queries and other user activity.

There are two data types in the Knowledge Graph: **entities**, which are real-world persons, places, or things; and **concepts**, which are abstract ideas or constructs. For instance, the Knowledge Graph would define *Formula One* as an entity that is associated with the 'auto racing' concept. *Williams Racing* is a Formula One team that is often referred to as simply 'Williams,' and is also connected to 'auto racing,' so it would be classified as an entity, and connected to the *Formula One* entity. *"World championship"* is an abstract concept which could apply to multiple entities, but is narrowed to the scope of Formula One, and then to the Williams Racing team, due to the co-occurrence of those entities in the search query.

You can see this process in real-time as you type; Google's **autocomplete** feature will show you what it thinks the next words in your query will be. The moment the words *williams F1* are typed into the query field, Google has already narrowed the search scope to the Formula One and Williams Racing entities and their relationships, and has calculated several words that could logically follow, most notably *world championships*. You can see another example showing entities and their relationships in Figure 2-?? (ch2-entities-and-relationships.png).

Vertical Indexes

As part of the crawling process, search engines discover and catalogue vertical content wherever possible. Each content type has a unique set of attributes and metadata, so it's logical to create niche indexes for them, rather than attempt to describe and rank them according to generic Web search criteria. It also makes sense to create a niche search engine for each vertical index, but it's conceivable that there could be scenarios where a vertical index would only be indirectly accessible through universal search.

Private Indexes

Over the past 20 years, Google has offered several different products for creating private search indices, and there are other third-party companies that offer similar products, as well as open-source search solutions. These can be implemented publicly for on-site search capabilities, or they can be implemented privately for intranet search for employees.

Most companies with an intranet will eventually need an effective method for its users to search its contents. In some instances, you may need to work with a hybrid search engine that returns results from both a private intranet and the public Internet. This may create a situation in which you will be asked to optimize pages that will never be on the public Internet.

The Search Engine Result Page (SERP)

The SERPs are dynamically rendered based on many different signals of user intent such as the search query itself, current trends / events, location, the user's device, user search history, and other user behavior data. Changing any one of those signals or conditions may trigger different algorithms for the same query and/or generate different search results for the same query, and some of these variations in results composition may offer better optimization opportunities than others (as we previously explained with vertical content).

Universal SERPS can vary wildly depending on the degree of confidence in interpreting the searcher's intent, the degree of confidence that certain results will completely satisfy that intent, and how accessible those results are to both the search engine and the searcher.

Organic Results

Organic search results are any results within a SERP which aren't paid ads, or results which are content owned and published exclusively by the search engine. These results can include vertical results and special features for pages that use **structured data elements** (on-page metadata that Google can use to construct SERP special features; this is covered in detail in Chapter 6).

One core component of the search results are listings that are pure text, showing URL links, a title for the result, and some descriptive text. Some of these may include some enhanced information about the result.

The Title and Snippet

The traditional organic search result included the page's literal <title> element, and either its <meta-description> element, or the first couple of sentences of page content. Today, both the title and the snippet are dynamically generated by Google, using the content in the <title> element, the main or visual headline shown on a page, heading (H1, H2, etc.) elements, other content prominently used on the page, anchor text on the page, and text within links that point to the page[2].

Organic search listings can be elevated to the top of the results as **featured snippets** for some queries, in some instances where Google has a high degree of certainty that the query can be answered by showing an excerpt of your page content on the SERP. Featured snippets generally contain longer excerpts than normal listings, and the URL for the web site from which the answer was sourced appears at the bottom instead of the top.

Cached and Similar Pages

Most organic results have a link to Google's most recent cached copy of the page, though some webmasters choose to opt out of this service for various reasons by utilizing the noarchive directive[3]. A cached copy of a page will generally show its text content in an unembellished fashion; there is no fancy styling, programmatic elements are either disabled or statically rendered, and images are either missing or slow to load. The typical use case for a cached page is to see the content when the site is unavailable due to a temporary service outage, but cached pages also give you some insight into what the Googlebot 'sees' when it crawls the page. If certain navigation elements aren't rendering, that could interfere with crawling.

Results for some sites also include a **Similar** link. This leads to a new SERP that shows a list of sites that Google recognizes as being closely related.

Special Features

Under a variety of conditions, both universal and local searches can generate SERPs with special features to highlight vertical and structured-data-driven content. Google adds new special features every so often, so we'll explain the ones that were current as of the printing of this book in the sections below, but we feel it's more important to stress that all future SERP special features will follow the same pattern of using structured data elements to display vertical results.

> ### NOTE
>
> Some special features (such as the map pack) generally increase search traffic; others (such as enriched results and featured snippets) may in some cases reduce traffic. Some SEOs choose to de-optimize a page when a special feature results in lost traffic, but before you commit to doing that, ensure that the lost traffic is worth fighting for. If your conversions didn't decrease, then the traffic lost to a SERP special feature was either worthless or unnecessary. For instance, if you're selling tickets to an event, and Google creates an enriched result for your sales page that enables searchers to buy tickets directly from the SERP, then you're likely to see more conversions and less traffic.

OneBox Results

When a query is straightforward and has a definitive answer that Google can provide without needing to source that information from a third party web site, Google puts a OneBox answer at the top of the SERP, followed by universal results. In some cases these are combined with a featured snippet in order to allow Google to provide supplemental information to the user. For example, these queries will reliably return OneBox answers regardless of your location, device, and other search signals:

- when is father's day?

- what time is it in Florence?

- what is the capital of Washington state? Figure 2-3 (ch2-what-is-the-capital-of-washington-state.jpg) show this query that combines a OneBox with a featured snippet.

Figure 2-3.

Knowledge Panels

When the query is straightforward but the answer is too nuanced or complex to deliver in a OneBox result, Google will put a Knowledge Panel on the right side of the SERP, alongside the universal results. Knowledge Panels are tables of common facts about a well-documented entity or concept from the Knowledge Graph. Sometimes they are generated based on structured data elements, but most often they are just a SERP-friendly repackaging of the Infobox element of the topic's Wikipedia page.

The Google Map Pack

In addition to filtering out results from irrelevant locales, search engines also generate special SERP features for queries that seem to have local intent. The most obvious location-specific feature is the map pack, which is a block of three local business results displayed under a map graphic that shows where they are. If you see a map pack in a SERP for any of your keywords, that's proof that Google detects local intent in your search query, even if you didn't explicitly define a location. You can see an example of a Map Pack (also referred to as the "Local 3-Pack") result in Figure 2-4 (ch2-example-map-pack.png).
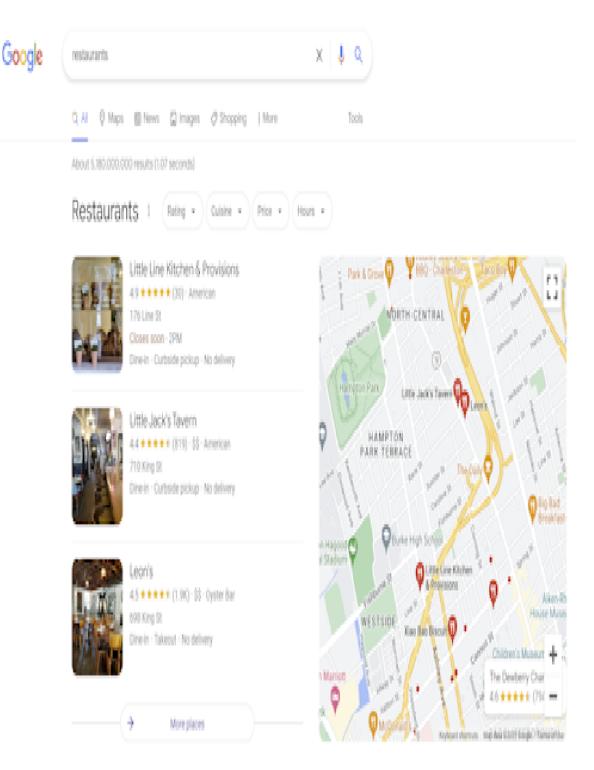
Figure 2-4.

For some verticals, the map pack shows business information on the left, and website and direction links on the right; for others, review scores are

displayed on the left, and images are shown on the right instead of links. On any map pack, you can click the "View all" or "More places" link at the bottom to see the Local Finder page, which generates a SERP focused entirely on local results. This is nearly identical to a vertical search on Google Maps, though the search radius will differ between the two depending on search intent. In addition, there will be variations in layout and functionality depending on whether you are searching on mobile or desktop, as will non-local search results.

Rich Results and Enriched Results

A rich result is when the snippet is enhanced or replaced by an image thumbnail or review star rating that summarizes or represents the originating content. Examples of when you'll find rich results include when you search for reviews of just about anything, or interviews with public figures. You can see an example of this in the screen shot for the search results for "eggplant recipes" shown in Figure 2-5 (ch2-eggplant-recipes.jpg).

Figure 2-5.

Searches for job postings, recipes, or event listings can generate enriched results. These go one step further than rich results by offering some kind of interaction with the page, sometimes without having to visit the page directly. For instance, you might be able to buy tickets to a show, track a package that has been shipped to you, or send in your resume to a recruiter directly from the SERP.

Carousel

If there are multiple relevant rich results, Google may show them in a carousel at the top of the SERP. You can see a good example of this if you query for: *starting lineup of the New York Yankees*.

This can also apply to universal queries that contain a high number of images, such as *wading birds*, or a high number of product results, such as *phones under $500*.

Key Moments in Video

Google is also likely to show enhanced results when videos may be a more useful result for users. This can include one or more videos, as well as links to specific moments within the video content. You can see an example of this in Figure 2-6 (ch2-how-to-build-a-deck.jpg)

Google

how to build a deck

About 900,000,000 results (0.84 seconds)

Videos

How To Build A Deck | DIY Home Improvement

YouTube · First Coast Craftsman

Nov 2, 2018

10 key moments in this video

From 00:14   From 01:26   From 00:01   From 04:04   From 06:47
Design the   Building   Pour concrete   How to   Laying a
deck using   permit   and set posts   calculate post   posts
Lowes Deck...   length

How To Build an 8 x10 Deck for Beginners

YouTube · Craig Heffernan

Jun 28, 2013

10 key moments in this video

Learn How To Build a Deck with Wood | DIY Projects

YouTube · Lowe's Home Improvement

Aug 3, 2018

10 key moments in this video

Feedback

Figure 2-6.

Sitelinks Searchbox

If Google determines that your query should have been executed on a specific site's internal search engine instead of Google, then you may see a search field below that site's snippet within the SERP. If you use that secondary search field for a query, then the search scope will be limited to that site (identical to using the **site:** operator).

To trigger a sitelinks searchbox, the site in question has to have its own publicly-accessible search feature, the Google search query has to be relatively broad, and it has to rank highly in the results. One example of this is if you search Google for *pinterest*, as you can see in Figure 2-7 (ch2-pinterest-seitelinks-searchbox.jpg).

Google   pinterest   X   🎤   🔍

Q All   📖 Books   🛒 Shopping   📰 News   🖼 Images | More   Tools

About 6,010,000,000 results (0.62 seconds)

https://www.pinterest.com   ⋮

Pinterest ✓

Discover recipes, home ideas, style inspiration and other ideas to try.

    Search pinterest.com   🔍

Login ✓

Forgot your password? Log in. No Account? Sign up. Are you a ...

Pins ✓

Pinterest | Waiting to be inspired? Welcome. That's what we're ...

Pinterest Business ✓

Pinterest is where people feel safe to explore new ideas and try ...

Help Center ✓

Ask us anything. The Pinterest Help Center is the place to get ...

Figure 2-7.

Query Refinements and Autocomplete

Google has invested heavily in machine learning technologies that specialize in analyzing human language – both spoken and written. Below is a list of some of those algorithms with descriptions of their purpose within Google Search:

The Synonyms System

While Google does not have an algorithm called the "Synonyms System" they do have algorithms that analyze keywords in a query and consider words and phrases which are similar in intent. This is highly similar to using the **OR** operator to list multiple similar keywords, except it's done automatically for you. For instance, if you searched for *Christmas gifts for kids*, a page optimized for *holiday presents for children* could rank highly in the SERP despite not including the literal keywords anywhere on the page.

This type of analysis also extends to how Google analyzes the content of web pages, and can extend beyond literal synonyms and to contextually understand the relationship between potential antonyms. For instance, a query for *where to sell a guitar* will likely include highly-ranked results for *where to buy a guitar*, because the concepts of *buying* and *selling* are very closely related in search intent, even if they are semantic antonyms.

BERT

Analyzes a sentence (or query) to determine the meaning of each of its words in context. Prior to BERT Google's language analysis could only consider the words before *or* after a word or a phrase to understand its meaning. BERT enables Google to examine words before *and* after a word or a phrase to fully understand its meaning. after For example, BERT enables Google to understand the intended meaning of *2022 brazil traveler to usa need a visa*, whereas prior to BERT Google would have assumed that the query was about someone in the US wanting to travel to Brazil (when in fact the opposite is the case).

SMITH

Analyzes sentences within the context of paragraphs or pages of content. For example, if you have a 2,000 word article about installing certain types of windows, and the article contains a unique piece of content about how to determine whether your windows contain UV glass, the SMITH algorithm can help extract that specific piece of content.

MUM

Uses language models similar to BERT and SMITH to answer questions that cross-reference two entities or concepts within the same topical domain. For instance, MUM could enable Google to assemble a SERP specific to the mixed context of a complex query such as: *I'm a marathon runner. What do I need to know to train for a triathlon?*

> **NOTE**
>
> As of this writing, Google is still testing MUM, and has not put it into production in universal search.

Search Settings, Filters, and Advanced Search

In addition to using search operators to limit the search scope for a query, you can use the **Advanced Search** feature buried in Google's SERP Settings menu, also located at this specific URL: https://www.google.com/advanced_search.

From that Settings menu (the gear icon on the SERP, or the Settings link in the lower right on the main search page), you can also alter Search Settings to limit results to pages that are hosted in certain countries or contain content in certain languages, change the number of results per page, enable or disable Autocomplete, and enable or disable the **SafeSearch filter** that attempts to exclude broadly 'offensive' results.

On the SERP, the **Tools** menu enables you to filter results by age or date range, and to enable or disable verbatim interpretation of the query.

Google Ads

Google Ads is Google's parent company Alphabet Inc.'s primary source of revenue, generating 80% of revenue for the company in 2021[4]. In general, paid search placement is not within the scope of SEO. However, Google Ads data can be useful for keyword research purposes, and for A/B testing of organic landing pages, and many SEO professionals use Google Ads data, including conversion and user engagement data, to assist in SEO strategy development.

Ranking Factors

Search results are selected and ranked according to various logical processes (algorithms) which apply a variety of scoring methodoliges and rulesets.

In the early days of Web search, search engines were not advanced in how they assessed the quality of site content; they simply matched document vocabulary and user vocabulary. Pages which contained a title, description, or content that matched the verbatim search query would often reliably rank well for some or all of the keywords in that query, even if those sites were low quality content, or worse-spam. Consequently it was pretty easy to influence search results by inserting keywords into pages in the right locations.

The magic that set Google apart from its early search competitors was in the way it qualified pages in the index by analyzing whether they were linked to, and how they were described by, other web pages. The **PageRank** algorithm uses human-authored link text (the blue underlined text between the <a> HTML tags on webpages) as an extra layer of descriptive metadata for sites, pages, keywords, and topics, then evaluates each page's rankings based on the quantity of those links and the quality of the source pages containing the links.

More concisely: *links are votes*, (but all votes are not equal, and some don't count at all). The details involved in weighting those votes are among the most critical secrets for Google, but there are some fundamental truths. In general Google gives more weight to links from sites that:

- Are trusted by Google

- Have existed for longer periods of time

- Generate user traffic

- Are linked to from other trustworthy, topically related web sites

- These factors are discussed in depth in the following subsections.

Relevance

The first and most important objective for a search engine, as described earlier, is to deliver results that are relevant to the query by satisfying the intent of the user performing that query.

When a trusted site uses descriptive link text to link out to a page, Google begins to establish topical relevance for that new page. For instance if your site sells used cars in the Phoenix, AZ region, a descriptive link like this one from the Phoenix Chamber of Commerce will establish your site's relevance to the *Phoenix, AZ* topic:

```
<a href="https://www.example.com/index.xhtml">Local used car
virtual showroom</a>
```

If that page is also linked to from a credible national or international magazine site that reviews online used car dealerships, then that will establish relevance to the *used cars* topics – even if the link text does not contain the phrase "used cars," because Google understands that a site that has been publishing reviews about used car dealerships for the past 15 years is within the *used cars* domain.

With those two links, your site will soon be included in results for the query: *used cars in Phoenix*.

The words, phrases, and subjects that define your site in human terms can be referred to as its **topical domain** (a topic covered in more depth in Chapter 5: Keyword Research). All other pages on all other sites that share your topical domain(s) are relevant to your site in some way; the more

topics that you have in common with an external page, the more relevance you share with it.

Inbound links that originate from highly relevant sources are generally more valuable, in isolation, than links from partially- or tangentially-related sites. Links from sites that have no relevance to yours (which don't share your topical domain at all) are generally less valuable, unless you're truly beginning your link building efforts from scratch with a domain that has no links pointing to it yet.

AI / Machine Learning Impact on Relevance

While content and linking are important factors in Google's determination of a piece of content's topic and relevance to a query, modern machine learning technologies such as Google's BERT, SMITH, MUM, and other algorithms are used to as well to analyze queries and the text content of webpages. The result is a set of AI models that can analyze new webpages and natural language queries, then determine their relevance to every topical domain in the index.

These algorithms, however, also have drawbacks which are difficult to address programmatically – including containing inherent human biases toward race, gender, and various other social and cultural elements, as well as being vulnerable to the influence of organized disinformation campaigns, the interpretation of negative associations, and the semantic collisions caused by slang terms and regional dialects. Generating a new AI model also requires a massive amount of computing resources, which uses a massive amount of energy at a company that claims to be committed to being carbon-neutral.

EAT

Regardless of algorithms, the SEO fundamentals remain unchanged: conduct good keyword research that includes natural language queries, build web content that is accessible to search engines, and create quality,

useful content that attracts high-quality links to your responsive, cleanly-structured website.

This first came to light when Google published a new version of their Search Quality Raters Guide (SQRG) in 2018. The full concept was known as EAT and this was an acronym for Expertise, Authority, and Trust. The concept of EAT as discussed in the SQRG is not a direct ranking factor but is instead a series of factors that human reviewers are asked to consider in evaluating the quality of the search results.

These are not direct ranking factors and the human reviewer's input is not fed directly into the Google algorithm. Instead, it's used to highlight examples of Google search results that need to be improved. This can then be used as test cases when Google engineers work on new algorithms to improve the overall results. EAT as well as the individual component signals (Expertise, Authority, and Trust) are discussed in more detail in Chapter 6. These are also more briefly introduced here.

Expertise

Expertise relates to the depth of knowledge that you offer on your site. For example, contrast the expertise of a general copywriter you hire to write your content with someone that has two decades of experience in the topic area of your business. A general copywriter given a few hours to create a piece of content will struggle to write the same quality of material that a true subject matter expert can.

Authority

Google assigns authority to sites that are linked to from other authoritative sites that have proven to be trustworthy over time. Your used car website will gain authority when it is linked to from external pages that have a significant amount of relevant topical authority in the search index.

While relevance is easy to determine, authority is not; calculating authority requires a nuanced interplay between many objective and subjective factors, and every search engine has its own methods for calculating authority.

Trust

A search engine's concept of trust is similar to the usual sociological definition: it's a measure of a page, site, or domain's integrity over time. Trusted sites have a long history of consistently playing by the rules and have not been compromised by spammers or scammers. If Google had to rebuild its index from scratch, one of the first things that information architects might need to do is create a list of trusted sites. From there, authority would be calculated as the product of relevance and trust. This is not conceptually different from the process of moving to a new town and establishing a new social network; you identify people you can trust, then expand outward from there because trustworthy people usually associate with other trustworthy people, but as you get further away from the original trusted source, you have to reduce the level of inherited trust. Using this theoretical approach, a site that is one click away from a highly-trusted source will inherit a lot of trust; two clicks away, a bit less; three clicks away, even less; and so forth.

Again, search engines consider their actual trust algorithms to be valuable trade secrets, so the best you can do is examine the results and try to work backwards toward the contributing factors, which is exactly what link analysis tools and SEO platforms do. While the technical details will always be obscured, it's safe to assume that search engines follow the same paradigm as a human "background check." Which sites does it link to, and which sites link to it? What's the history of this domain, including hosting and ownership? Does it have a privacy policy and valid contact information for the owners? Is it doing anything suspicious with scripts or redirects?

Local Signals and Personalization

As explained earlier in this chapter, results can be heavily influenced by local intent and personalization factors – essentially, taking user behavior data into account when determining query intent and deciding which results to show. For example, depending on the query, searches from mobile devices may be assumed to be 'local first.' The stronger the local signals, the less likely it is for non-local results to appear in a SERP, even if they are

highly relevant to the query. For instance, if you were to execute this query from a desktop computer with an IP address originating in Cleveland, Ohio, there would likely be few or no local results:

Disney's haunted mansion

This is the title of a movie, and the name of a theme park ride at both Disneyland in California, and Walt Disney World in Florida. If there are no local signals, the SERP will mostly pertain to the movie. However, if you were to execute the same query from a smartphone with an IP address indicating the Orlando, Florida area, the SERP would be more likely skew toward the theme park ride at Walt Disney World, and any results for the Disneyland ride would be more likely to rank considerably lower no matter how highly they rank for the *Disney's haunted mansion* keyword. You could arrive at a similar SERP from your Ohio-based desktop computer just by adding *Florida* to the query:

Disney's haunted mansion Florida

You'll still get a few results (and probably ads) pertaining to the movie, since there is likely to be a shared interest between the movie and the theme park ride. However, if you were to execute these queries in reverse order (with *Florida*, then without), the second SERP would be more likely to be nearly identical to the first. Why? Because you recently expressed interest in the theme park ride in Florida, so Google may assume that this is still your intent even though you removed *Florida* from the query. By taking these factors into account, Google is what is generally referred to as a "biased" search engine.

The impact of personalization can extend well beyond your most recent searches. If you're signed into a Google account and allow Google to collect historical data about you, many SERPs you see will be impacted by that data. Search history is not limited to keywords; it also reflects vertical search habits. For instance if you often search for videos from your smartphone, but rarely from your desktop computer, then your mobile SERPs will be more likely to contain videos.

Timing and Tenure

Search engines keep detailed records on linking relationships between websites (as well as information pertaining to domain names, IP addresses, sites, pages, and URLs). With regard to linking relationships, the search engines generally store:

When did the link first appear?

This isn't just a simple datestamp; it's combined with an analysis of other changes in the index. Did this link (URL) appear immediately after an article was published in the *New York Times?*

When did the link disappear?

Sometimes link retirement is routine, such as when blog posts move from the home page to an archive page after a certain period of time. However, if an inbound link disappears shortly after you made major changes to your site, search engines may interpret this as a negative signal. Did that site's owner disagree with the changes you made, and revoke its association with your page?

How long has the link existed ?

If a link has been around for a long time, a search engine can potentially give it more weight or less, depending on the authority/trust of the site providing the link, or other secret factors.

How quickly were the links added (also known as "link velocity ")?

Drastic changes in the rate of link acquisition could also be a signal of something good or bad from a search engine's perspective. For example, if your site is featured in major news coverage, it could be good. If you start buying links by the thousands, it would be bad.

Legitimacy

Google analyzes the context around links to determine their legitimacy. In a previous era, search engines were fooled by keyword-stuffing and link-

farming. Google, in particular, now goes to great lengths to detect link schemes and spammy content, and also to explicitly detect legitimate content. Here are some of the potential factors that search engines may use to qualify content:

External links to the linking page

Does the external page containing the inbound link have its own inbound links? If the page linking to your site is benefiting from incoming links, then this will make the link to your site more valuable.

Nearby links

Do the closest links on the page point to closely related, high-quality sites? That would be a positive signal to the engines, as your site could be seen as high-quality by association. Alternatively, if the two links before yours are for unrelated sites that may be low quality content, this can possibly reduce the value of your link.

Page placement

Is your link in the main body of the content? Or is it off in a block of links at the bottom of the right rail of the web page? Better page placement can be a ranking factor. This is also referred to as *prominence*, and it applies in on-page keyword location as well.

Nearby text

Does the text immediately preceding and following your link seem related to the anchor text of the link and the content of the page on your site that it links to? If so, that could be an additional positive signal. This is also referred to as *proximity*.

Closest section header

Search engines can also look more deeply at the context of the section of the page where your link resides. This can be the nearest header tag,

or the nearest text highlighted in bold, particularly if it is implemented like a header (two to four boldface words in a paragraph by themselves).

Overall page context

The relevance and context of the linking page are also factors in the value of a link. If your anchor text, surrounding text, and the nearest header are all related, that's good. If the overall context of the linking page is also closely related, that's better still.

Overall site context

Another signal is the context of the entire site that links to you (or perhaps even just the section of the site that links to you). For example, if a site has hundreds of pages that are relevant to your topic and links to you from a relevant page, with relevant headers, nearby text, and anchor text, these all add to the impact, so the link will have more influence than if the site had only one page relevant to your content.

Source Diversity

In addition to sourcing links from similar web site types, you should also try to get links from pages that have different content and serve different purposes.

For example, if all your links come from blogs, then you have poor source diversity. There are many examples of other types of link sources: national media websites, local media websites, sites that are relevant but cover more than just your space, university sites with related degree programs, and so on.

If all your links come from a single class of sites (e.g., blogs), search engines may view this as a potential link scheme. If you have links coming in from multiple types of sources, search engines are more likely view your backlink profile as legitimate.

Keywords in Anchor Text

Anchor text (also called *link text)* refers to the clickable part of a link from one Web page to another:

```
<a href="http://www.example.com">This is anchor (or link) text.</a>
```

Search engines use anchor text as descriptive metadata about the destination page. However, don't try to stuff keywords into anchor text if the words don't naturally fit with the surrounding content, and avoid overly descriptive anchor text can appear to be keyword stuffing. Search engines look for unnatural language usage in anchor text, and if they detect this, they will lower the ranking of the linked page. Similarly, if you have 20 external links to your page and 19 of them use anchor text that matches your main keyword exactly that can be seen as unnatural as well, and these links may be discounted.

Negative Ranking Factors

It's also possible to have negative ranking factors. For example, if a site has a large number of low-quality inbound links that appear to be the result of artificial efforts by the publisher to influence search rankings, the links can be ignore, and in extreme cases the site's rankings can be lowered. This is, in fact, exactly what Google's Penguin algorithm does. This algorithm is discussed more in Chapter 9. Some other potential negative ranking factors include:

Malware hosting:

Your site must not contain malicious software or scripts. Usually this happens by accident – your site is hacked without your knowledge, and malware is hosted clandestinely.

Cloaking :

Your site must show the same content to users that it shows to search engines. If you try to circumvent this by showing a special page to Web crawlers, your site will be penalized. Please note that with Client Side Rendering (CSR) there can be scenarios where users see different

content than search engines, and this is typically not seen as cloaking. This is discussed more in Chapter 6.

Unqualified paid links:

If you sell links from your site to others, they must be properly marked with an appropriate link rel=sponsored attribute (see Chapter 6 or Chapter 10 for more details), or else your site could potentially be penalized.

Page load time:

If your site's content is very slow to load, its visibility within search results can be negatively impacted, as the search engines are aware that users generally seek to avoid pages that don't load quickly. Fix your site first, then optimize it for search.

Link acquisition rate (link velocity):

Search engines are conscious of how a site's search visibility changes over time. If your site has acquired an average of 5 links per day, and then the links suddenly start to come in at a rate of 10 per day, that could be seen as a positive ranking signal. If the rate of new links drops to 2 per day, that could be a signal that your site has become less relevant. If your site suddenly starts to get 300 new links per day, then it has either become a lot more relevant (in which case there should be a lot of social media mentions and news stories), or you're gaming the system in some way (in which case your site will eventually be caught and penalized).

User Behavior Data

Search engines – Google in particular – monitor user behavior data, including their interaction with SERPs, with the goal of providing a better search experience to ensure repeat users. This data includes (but is not limited to) location data, voice search data, mouse movements, and data

gleaned from authenticated Google Account usage across various Google products.

While click-throughs from search results are a good signal for both search engines and web sites, if visiting users quickly abandon your page and come back to the SERP (a **bounce**), this can be a negative signal. Users can bounce back to the SERP for a variety of reasons, including slow page load time, poor user interface design, irrelevant content, being presented with interstitial ads or paywalls, or simply by accidentally clicking or tapping the wrong result. Generally speaking, a high bounce rate can be a signal that something is wrong with your site or that you're providing a negative experience for users.

Conclusion

Although search technology and the digital ecosystem as a whole continue to rapidly evolve, having a solid understanding of search engine fundamentals is absolutely essential to SEO success. Understanding how various signals are used by search engines to deliver a good search experience for users to ensure their return, you're in a good position to develop a strategy to connect your future customers to the content you create in order to leverage organic search for your business. Next, we will evaluate and assemble a set of tools that will help you collect and analyze data about your site and the search context that will most effectively connect people to it.

---

1 https://blog.google/products/search/introducing-mum/

2 https://developers.google.com/search/docs/advanced/appearance/title-link

3 https://developers.google.com/search/docs/advanced/robots/robots_meta_tag#noarchive

4 https://abc.xyz/investor/static/pdf/20210203_alphabet_10K.pdf?cache=b44182d

CHAPTER THREE

Your SEO Toolbox

Success in SEO is highly dependent on having the right tools. Before you can learn the tricks of the trade, you need to have the tools of the trade. The services and utilities covered in this chapter will enable you to analyze your site and identify technical or structural problems, discover the most cost-effective topics and keywords, compare your site's performance to its top competitors, track incoming links, and measure visitor behavior. You're probably already familiar with some of them, but we're going to reintroduce you to them from an SEO perspective. They all have a variety of purposes and functions that are useful for SEO, but their common thread for most of them is their utility in keyword research, a topic that we cover in great depth in chapter 5.

The Artist's Tools

The first and most important tool is a simple spreadsheet application to aggregate data from multiple sources and calculate the best opportunities. This is a requirement for maximum SEO productivity and efficiency.

Next, we'll help you explore some options for technical SEO. It doesn't make sense to optimize a site that isn't being indexed properly. There are many technical utilities and site analysis features of larger SEO service packages that can help you solve technical site problems. Some are standalone (like Google Search Console), but every major SEO platform and most marketing services suites have their own site and page analysis features.

There are three data perspectives on website activity: server-side, client-side (visitor-side), and search-side. Since your company or client controls the Web server, it makes sense to start this chapter by first analyzing the data you already have. Next, we'll explain how you can supplement that data with extra visitor context from on-page JavaScript trackers. Finally, we'll introduce you to SEO platforms that provide search data for the keywords you're targeting, the current search rank for every indexed page on your site, and other valuable features that will help you optimize your site and increase search traffic.

Some of the tools we cover in this chapter are free, but most require a license fee, paid subscription plan, or SaaS contract. Paid tools tend to charge on a per-user, per-site (or property), or per-client basis. While you don't have to make any decisions about any of these tools right now, in order to follow many of the processes and examples throughout the rest of this book, you must have at least one SEO platform subscription and a Web analytics service or tag manager deployed on your site, and you must set up and configure Google Search Console.

Spreadsheets

Our dream careers rarely align well with reality. Real archaeologists spend a lot of their days on their knees in the hot sun with a toothbrush and a garden shovel, not dodging ancient booby-traps with a whip and a revolver à la Indiana Jones. Real lawyers spend much of their time on administrative tasks such as billing, collections, office management, case law research, and reading and filing formal documents, not winning clever courtroom battles with hostile witnesses or delivering dramatic closing arguments before a jury. And a professional SEO often spends more billable time in a spreadsheet than a Web browser or code editor. Hopefully that doesn't throw too much water on your fire. This is still a fun and fascinating industry!

SEO practitioners rely heavily on spreadsheets, and most commonly that means Microsoft Excel, but you can use any modern equivalent. Regardless of which spreadsheet app you use, you must be proficient enough with it to create and maintain proper keyword plans for your company or clients. Specifically, you must be comfortable working with data tables, basic formulas, filters, and pivot tables. If you have a few knowledge gaps in these areas or you don't feel confident in your spreadsheet skills, then you should invest in training, or at least be prepared to use the Help menu and Google to figure out how to use these advanced features to filter, sort, and calculate your keyword lists.

This topic is covered in more detail in Chapter 5, where we walk you through the process of creating a keyword plan spreadsheet.

Traffic Analysis and Telemetry

In order to analyze data, first you must collect it. There are two paradigms for visitor data collection: raw Web server logs that record all incoming traffic from the Internet, and JavaScript trackers (also known as tags) that are embedded in the source code of every page on your site. Each has its advantages and disadvantages, so the ideal solution would incorporate both Web server log input and JavaScript tracking to give you a holistic perspective on traffic and user behavior, but this depends on your individual

feature requirements, and what's already deployed (or required due to vendor contracts) at your company.

Before you proceed, be aware that there is almost certainly already some kind of Web analytics package deployed somewhere at your company – perhaps several. It isn't unusual for a Web-based business to use a variety of separate analytics tools to measure different metrics or supply data to different services or utilities. You should begin by taking stock of what's already deployed (and paid-for) before making any decisions on analytics tools. As a consultant, there will also be times when you're stuck using a client's preferred vendor or solution, so you'll have to learn to work within those boundaries.

If this is a green field (meaning, nothing is currently deployed), then free analytics packages such as Google Analytics and Open Web Analytics are an excellent starting point. Even if a free service doesn't ultimately meet your needs, at the very least you can use it as a basis for comparing against paid alternatives.

Whether you're evaluating an existing solution or a new tool, make note of the features that you find valuable, and gaps in functionality that a competing program might be able to cover. Then look for ways to modify or extend the services you're using, or for a higher-end solution that covers those gaps. This is a long journey, not a one-time event. As you gain more experience in SEO, you'll continue to develop your requirements and preferences, and will likely end up with a few different go-to options for different scenarios.

Be wary of services (especially free ones) that want you to upload customer data or server logfiles; those service providers may collect your data for other purposes, and this would probably represent a privacy and/or security violation at your company. JavaScript trackers may also share or collect data about your traffic and visitors, and though this is less invasive and dangerous, it still may violate privacy laws or your internal IT policies. (Legal and privacy issues are covered in more detail in Chapter 13).

Google Search Console

https://search.google.com/search-console

This is a free service that provides a lot of technical site information that Google Analytics lacks. With it, you can test your site for indexation, view inbound search query data (keywords, impressions, clickthrough rate, rank), generate and test XML sitemaps, test mobile compatibility, analyze page performance, and measure the performance of structured data elements that generate SERP features such as enriched results and OneBox answers.

Google Search Console should almost always be the first service you configure for an initial site audit because it offers a quick way to identify low-level problems. Its keyword research utility is limited to SERP impressions your site is already receiving, but the data on existing search traffic is useful.

Server-Side Log Analysis

Web server software outputs a constant stream of text that describes all of the activity it is handling and stores it in a file somewhere on the server. A typical server log is a plain text list of HTTP requests. Here's an example of a line that you might see in an Apache Web server log:

```
127.0.0.1 [05/Nov/2015:21:43:06 -0700] "GET /requested_page.xhtml
HTTP/1.0" 200 1585 "http://www.example.com/referring_page.xhtml"
"Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:36.0) Gecko/20100101
Firefox/36.0"
```

From left to right, here's what that data represents:

1. The IP address of the machine (computer, mobile device, or server) that made the request.

2. The time, date, and timezone of the request (relative to the Web server).

3. The HTTP request method (either GET or POST) and the resource being requested (in this example it's a Web page named requested_page.xhtml).

4. The HTTP status code (200 represents a successful request).

5. The size of the request in bytes (usually either the amount of data being returned to the client, or the size of the file being requested).

6. The full URL of the page that referred to this resource (sometimes there isn't a referrer, such as when someone directly types or pastes a URL into a browser; referrers are only shown if the request came from an HTTP resource, such as when a user or crawler follows a link on a Web page).

7. The useragent string, which shows the browser or crawler name and version number (Firefox, for some reason, likes to report itself as "Mozilla"); the operating system name (X11 is a graphical user environment framework on the Ubuntu Linux operating system), revision or build number (expressed here as rv:36.0), and CPU architecture (x86_64 refers to an Intel or AMD 64-bit processor); and the HTML rendering engine and revision number (Gecko is the native engine in the Firefox browser, and the revision number is the same as the operating system's in this example because it was built for and distributed with Ubuntu Linux).

NOTE

If you'd like to learn more about the elements of an HTTP request and how to adjust the verbosity and format of a log file, consult your server software documentation.

Everything about these log files is configurable, from the level of detail and verbosity to the location of the file and the conditions under which one log file should close and another one should begin. The Web server log is the oldest tool in the box, and when it is properly configured and its data is filtered and aggregated for human consumption, it can be useful for a variety of SEO purposes:

- How often search engines are crawling the site. More importantly, how often they crawl each of your pages, including those that they don't crawl at all.

- How much time is being spent crawling low value pages. This is time that they spend crawling pages that you should prefer that they would spend that time crawling more important pages on your site.

- Identifying pages that redirect using a means other than a 301 redirect.

- Identifying chains of redirects

- Pages on your site which return status codes other than "200 OK"

- Backlink discovery

- Finding missing pages / bad links

- Site performance

- Visitor platforms (device, operating system, and browser version)

- Visitor locale

Web server logs can also be merged with other data sources to provide insights on conversion rates from paid and organic campaigns, server optimization, and URL canonicalization for duplicate content, among other things.

NOTE

Some Web hosting providers may restrict or deny access to raw server logs and configuration files. Web server logs are useful for technical SEO, so if the hosting company won't give you the data you need, you may want to switch to a more SEO-friendly provider.

The raw log file from a Web server, while human-readable, isn't human-comprehendible. You need a third-party analysis tool to cleanse, combine, sort, and slice this transactional information into facts and dimensions that actually mean something to you. There are so many different tools and methods for logfile analysis, it could take up a whole (rather boring and repetitive) chapter of this book, so in the subsections below we just mention a few.

First – as always – check to see if there's already something deployed, and whether it will meet your needs. Keep in mind: a server log is nothing more than a plain-text transactional data source. Aside from dedicated Web logfile analysis tools, many companies have business analytics packages that can use server logs to produce useful reports and dashboards. A business analyst within the organization can potentially work with you to pull Web log data into it and configure the output to meet your needs. Also, check to see if the company is already using an enterprise services suite (such as Atlassian, Salesforce, IBM, or Oracle) that has an optional Web log analysis component. Even if the company's not currently paying for that component, you'll have a much easier time convincing management to expand the budget for an existing solution than to buy into a completely new one.

The subsections below cover some logfile analysis tools that we've used and would recommend.

Splunk

https://www.splunk.com

Splunk bills itself as a "data to everything platform," meaning it can be configured to use any logfile or database as input, and produce any kind of report, chart, or file as output, which means that Splunk isn't functionally different from most other business intelligence or data analytics solutions. It may be a more affordable alternative to larger analytics suites, though, because Splunk is more of an engine than a service, meaning you have to develop your own solution with it rather than just copy-and-pasting some code or clicking a few buttons in a Web interface ala Google Analytics.

Splunk is an extremely high-end option, and may be overkill if you're only analyzing relatively small transactional Web logfiles, but it's popular enough that it may already be deployed elsewhere within a large organization, which would make it cheaper and easier for you to adopt for your SEO project.

DataDog

https://www.datadoghq.com

Whereas Splunk is more of an analysis tool, DataDog is more of a real time monitoring service. Like Splunk, it can also take input from several different logfiles, but its output is geared more toward real time results. This is a good solution for measuring the efficacy of time-limited campaigns, short-term promotions, and multivariate testing.

The Screaming Frog SEO Log File Analyser

https://www.screamingfrog.co.uk

This may be the perfect standalone tool for Web logfile analysis for SEO projects. It's inexpensive, and its only purpose is SEO-oriented Web server analytics, so you aren't paying for features and functions that have nothing to do with your work.

If you don't want to buy into a long-term agreement with a service provider, but need solid server-side analytics, Screaming Frog should be your first consideration. The free version has all the same features as the paid version, but it's limited to importing only 1000 log lines, so it's more of an "evaluation edition" than a viable free edition – you'll be able to see if it has the features you need, but with such a big limitation on the amount of input, it isn't viable in production the way other free solutions are (such as Google Analytics and Open Web Analytics).

OnCrawl

https://oncrawl.com

This is a full-service SEO platform that has particularly good SEO-oriented site analysis tools which combine crawler-based performance metrics with logfile data. You don't have to subscribe to the full suite – you can just pay for the logfile analyzer – but its other SEO components are worth considering.

Botify

https://botify.com

Botify is a large-scale technical SEO analysis service that can use multiple local and external data sources to provide insights and metrics on all of your digital assets. As with OnCrawl, logfile analysis is just part of its larger toolset, and it's worth your while to evaluate it in a larger technical SEO context.

JavaScript Trackers

Web server log stats are a valuable SEO asset, but they can only show you what the server can record. From this perspective, it can sometimes pose a challenge to tell the difference between a real visitor and a bot using a false user string, and you'll have little or no ability to analyze user activity across multiple devices and sessions. While most search crawlers identify themselves as bots via unique user agents such as Googlebot, DuckDuckBot, and ia_archiver, anyone can write a simple program to scrape Web pages and use whatever user agent they like, and your server will record whatever they report themselves as. A badly behaved bot can circumvent a site's robots.txt restrictions and pollute your server log with bad data (not to mention overloading your server and network). Note though that you can still block these bots with .htaccess directives or a service like IPBlock.com. Most of the time this isn't a malicious act – it's just somebody clumsily trying to understand what's on your site.

JavaScript trackers work at the browser level, so they naturally weed out crawlers and bots because they aren't proper browsers and to save on resources most of them don't execute JavaScript. If they do, it's on a limited

basis strictly to evaluate content on a dynamically-generated page, which does not typically ping a JavaScript-based analytics service. For example, Googlebot is a search engine crawler that executes JavaScript.

JavaScript trackers also offer more in-depth visitor information such as time on site, clickstream data (the exact mouse clicks or screen taps a visitor makes when navigating a site), and demographics (if visitor data is being stored in a cookie that your analytics package has access to). However, many people use VPNs, browser plug-ins, and other privacy countermeasures that prevent JavaScript trackers from working properly (though that data will still show up in your server logs). Assuming that your trackers are not blocked they'll provide you with a highly detailed view of a subset of the true visitor data. Over time, this method will become less effective as privacy and security awareness increases. For that reason, many modern solutions attempt to show you the whole picture by pulling in external data from server logs or other tracking services, and integrating it with what's been collected via JavaScript.

Google Marketing Platform

https://marketingplatform.google.com

This is Google's suite of Web analytics products, which encompasses the following components:

Google Analytics:

> a Web traffic and visitor behavior data collection and analysis tool. Google Analytics is the key component in this platform because it's the primary data collector. You can use it standalone without involving other aspects of the Google Marketing Platform. Refer to the Google Analytics section later in this section for more details.

Looker Studio (formerly known as Data Studio):

> a basic data analysis tool that enables you to combine data from several sources, and create reports and dashboards to visualize certain metrics and goals. This is not nearly as powerful as a dedicated business

intelligence solution such as Cognos, Qlik, or Oracle, but you may be able to make it work for your project.

Tag Manager:

a framework for deploying multiple JavaScript trackers using one lightweight tag. Refer to the Google Tag Manager section later in this chapter for more details.

Optimize:

an extension of Google Analytics that enables you to perform A/B testing on content pages, and to deploy custom segmentation rules for certain groups of visitors.

Surveys:

an easy-to-use utility that helps you create and deploy visitor surveys on your site.

Everything in the Google Marketing Platform is free, but all data recorded through these services will be collected and used by Google. There is of course a paid upgrade to the enterprise service level which enables you to keep your data private, adds a few new services that help you track marketing campaigns and manage digital media assets, and includes "360" branded editions of the free services listed above. The 360 editions include Campaign Manager 360, Display & Video 360, and Search Ads 360. These have more integrations with other services, more filtering and funneling options, and higher limits on Web property views and data caps, among other perks and upgrades.

In this chapter we're only covering Google Analytics and Google Tag Manager. The rest of the Google Marketing Platform services may be useful to you for other marketing purposes, but they have little to do with SEO, or are inferior to the equivalent components in proper SEO platforms and business analytics suites.

Google Analytics

https://analytics.google.com

As shown in Figure 3-1, Google Analytics is active on more than 72% of websites (Google Universal Analytics and Google Site Tag Manager incorporate Google Analytics). The basic version of Google Analytics is free of charge, easy to deploy, and offers a deep view of visitor data. If you're an in-house SEO and only manage one site, and you don't have any security or privacy concerns with Google retaining the data it collects from your Analytics deployment (Google has access to much of this data anyway via its Googlebot crawler, cookie trackers, Chrome users, and visitors who are logged into a Google account during their browsing session), then the free version might be the perfect solution for you.
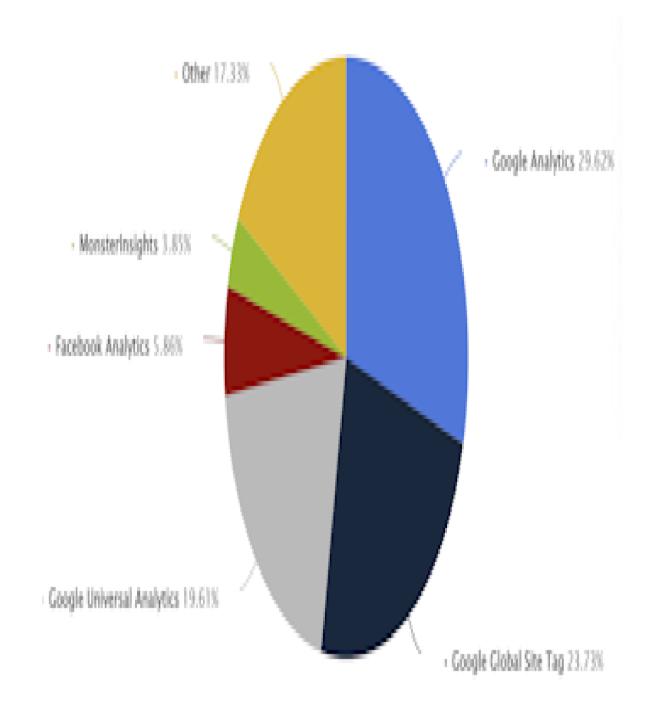
Other 17.33%

Google Analytics 29.62%

MonsterInsights 3.85%

Facebook Analytics 5.86%

Google Global Site Tag 23.73%

Google Universal Analytics 19.61%

Figure 3-1. Google Analytics market share (Statista)

If you have some concerns with Google's data collection practices, or if you're a freelance SEO and have multiple Web properties to monitor, then the upgrade to Analytics 360 is worth evaluating.

As with all individual pieces of larger service suites, when you're using other Google products like Google Ads, Google Search Console, or Google Ad Manager, the unique integrations that they have with Google Analytics may prove to be a major advantage over other analytics tools.

Adding "Not Provided" Keywords to Google Analytics

Google Analytics and Google Search Console usually don't reveal the keywords that people search for when they click through to your pages from a SERP. If you see the phrase "keyword not provided" instead of a search term in your analytics, then you know that Google has intentionally stripped the keyword data. To get the most SEO value out of Google Analytics, you'll have to use a third-party tool to add that missing information, such as:

Keyword Hero

> (https://keyword-hero.com) This tool was specifically designed to solve the "keyword not provided" issue in Google Analytics. It pulls your data from Google Analytics and Search Console services via APIs, uses machine learning to match each page with a list of likely search queries, then exports everything back to Google Analytics.

NOTE

Some SEO platforms (covered later in this chapter) can provide missing keyword data, too, most notably: Semrush, Ahrefs, and Searchmetrics.

Kissmetrics

https://www.kissmetrics.io/

Traffic analysis is all Kissmetrics does, so if you may want to give it extra consideration if you prefer to build your SEO toolset with interchangeable standalone services, or if you're unhappy with the Web analytics capabilities of a larger marketing suite that is already deployed at the company and are looking for a one-off replacement.

There are two different Kissmetrics analytics products: one for SaaS sites, and one specialized for ecommerce sites. Both services attempt to identify individual visitors and build a profile for them that encompasses all of their actions during all of their visits across all of their devices and browsers. This has obvious SEO advantages, but it could also be useful as a data source for expanding the information in your CRM.

Adobe Analytics

https://business.adobe.com/products/analytics.xhtml

According to Builtwith.com Adobe Analytics is used on more than 200,000 web sites. Adobe offers a comprehensive suite of online marketing tools in its Adobe Experience Cloud. Adobe Analytics (either separately or as a part of Adobe Experience Cloud) is used by a large number of large enterprises, due to the scalability of the platform.

As a standalone solution, Adobe Analytics has interesting features that most of its competitors don't. These include access to a wider array of potential data sources beyond what is collected via its JavaScript tag, and an AI component that can predict future traffic levels based on patterns and anomalies in past visitor data. It also is highly flexible and customizable and integrates with other data platforms.

Open Web Analytics

http://www.openwebanalytics.com

This is an open-source Web analytics package written in PHP. It isn't hosted externally – you have to deploy it on your Web server and configure it yourself (if you're not in a small business your system administrator would do that for you). The advantages are that you control your data and your deployment in-house; you can pare it down to just the data and visualizations that you need; and you have the unique ability to track via both PHP and JavaScript, which allows you to collect some data about people who use browser plugins or VPNs to block trackers, or have JavaScript disabled.

When properly deployed, Open Web Analytics looks and acts a lot like Google Analytics without all of its extra marketing metrics and Google product integrations. It's an excellent barebones Web analytics package with no bloat, no monthly fees (but if you use it in production, you should give back to the project via a donation), and no limit on how many sites you can deploy it to. The downside is, someone at your company must spend time and effort to deploy and maintain it.

Tag Managers

Your company or client may have several different JavaScript trackers deployed on the same site for various reasons. Multiple trackers will increase page load time and potentially interfere with one another. Tag management services combine several different JavaScript trackers into one short tag, which reduces page load time, bandwidth usage, and the effort of managing multiple tags on a large site.

Ultimately it's still better to reduce the number of tags per page, but in a large company that may not be possible due to interdepartmental politics and budgeting.

Google Tag Manager

https://tagmanager.google.com

Because it's free, this is an extremely popular tag management solution. The only downside is that it may not support all but the most commonly-used JavaScript trackers, which forces you to use custom tags for unsupported tags. That's only a deal breaker if you aren't savvy enough with JavaScript (or have access to someone who is) to create and maintain that custom code.

Google Tag Manager is undeniably the first solution you should explore, not just because it's free, but because it might be the simplest option.

Tealium

https://tealium.com

Tealium is a commercial tag manager that supports substantially more tracker code templates than Google Tag Manager, which makes it easier for non-developers to deploy. However, it may incorrectly render your pages for visitors who block all trackers, and ultimately it's the conversions that matter, not the visitor stats.

Search Engine Tools and Features

We've already covered a few Google services that are designed specifically for website optimization, but if you approach many of Google's other properties from a creative angle, you'll find a lot of hidden SEO value – particularly for keyword discovery and valuation.

Autocomplete

Google uses artificial intelligence to predict the rest of your query when you start typing in the search field, and shows you a list of top-ranked predictions below it. This is an excellent way to see the most popular keywords for your topics. For example, typing in *lilac* might reveal suggestions like those shown in Figure 3-2:

lilac

Lilac
Plant

Lilac Hedge Farm
Farm · Rutland, MA

lilac **blossom**

lilac **cookies**

lilac **blossom nashua nh**

lilac **sugar cookies**

lilac **diesel strain**

lilac **color**

lilac city grill
Lilac City Grille · Rochester, NH

Lilac House
Chinese restaurant · 34 Milliston Rd, Millis, MA

Report inappropriate predictions

Q All   News

About 813,000 re

Metrics Enterp

for acquiring o

websites or oth

https://www.kiss

Privacy - Kis

https://www.kiss

Google Chro

Figure 3-2. Example Google Autocomplete results

Google won't tell you how many times *lilac sugar cookies* has been searched for, but because it appears at the top of the list of suggestions, you can infer that it was probably searched for more often than the phrases that appear below it. This can give you important insight into what searchers are looking for, or what they search for in relation to a specific topic.

> **NOTE**
>
> Autocomplete predictions are strongly influenced by the user's location (e.g., *wet n wild phoenix* might be a prediction that shows up when a user in Phoenix, Arizona, types the letter *w* into Google.) If you're searching from a mobile device, you might also see apps in the list of suggestions.

Google Ads Keyword Planner

https://ads.google.com/home/tools/keyword-planner

The Google Ads Keyword Planner can analyze your site content and deliver a list of relevant keywords that are likely to be converted to clicks in a Google Ads campaign. It's free to use, but you have to have a Google account (which is also free) in order to log in. It has excellent documentation and a walkthrough for new users.

Regardless of whether your company invests in Google advertising, you can use the Keyword Planner to get good suggestions for related terms, search volume estimates, search trends, and ad cost estimates for any keyword or URL that you enter. You could use these numbers to calculate keyword popularity, keyword difficulty, and CPC (this process is covered in depth in Chapter 5). Unfortunately, unless you are running a paid ad campaign, the search volume will be approximated in wide ranges, which isn't precise enough for you to use for anything other than quick, basic keyword valuation. You can work around this and get better data by setting up a small, low-budget campaign; or if you know someone who is spending

money on Google Ads campaigns, you can ask them to add you as an authorized user to their account.

The CPC data is much more precise than the other metrics, and can be useful for gauging keyword difficulty for both paid listings and (in a general way) organic searches. You can get more exact estimates by selecting specific budgets or costs per click, and you can forecast the traffic impact and conversion rate based on data from the past two weeks.

The Google Ads Keyword Planner is most useful for projects that will include advertising, but it can also serve as a free (or low-cost) option for organic keyword research. Due to its limitations, though, this is no replacement for search data from any of the major SEO platforms (covered in the next section).

Google Trends

https://trends.google.com/trends

Google Trends enables you to view the relative popularity of a keyword over time and by geography. You can also compare trend data between two or more keywords. There are two datasets you can use: realtime (search data from the past 7 days, up to the past hour), and non-realtime (search data spanning the entire Google archive, starting in 2004 and ending about 36 hours ago). Both datasets are sampled (a random, representative sample is taken from the complete source, similar to a poll), normalized according to time and locale, and indexed to the timespan (on the line graph, the zero point represents the lowest search volume over that time span, and the 100 point represents peak search volume).

Google Trends is useful for identifying and predicting spikes in search volume for your topics, and for finding potentially related topics and queries. By drilling down into spikes in the trend graph, you can see the events and news stories that contributed to it, and that could give you some good ideas for site content and media outreach.

There's more you can do with Google Trends, including comparing relative volumes of different keywords (what's bigger, car repair or auto repair?) over time of multiple keywords, competitive brand strength research and seeing trends isolated down to any country in the world. You can read more about that in Chapter 5.

Google Trends is free to access (with a free Google account) and easy to use, so there's no harm or risk in looking up some of your topics or keywords to see if there's any useful information on them. Only popular search phrases are sampled, though, so obscure long-tail keywords often don't work with it.

Google News

Google News (and other searchable news aggregators) enables you to gauge the media activity around a topic, and the recent and real-time popularity (and therefore competitiveness / difficulty) of keywords. You can also search for mentions of your brand or company name (in quotes if it's a phrase). Results are sortable by time, all the way down to stories from the past hour, which makes Google News an excellent resource for extremely down-to-the-minute current keywords. The data from most other keyword research tools is at least a day old.

Related

Most search engines show "related" search results, either on the SERP or somewhere on a content page. Google shows related searches at the bottom of every SERP, and related queries under the "people also ask" heading. YouTube shows related videos in a sidebar next to the viewing pane. Twitter shows accounts that are similar to the one you're currently looking at. Amazon has several different "related" features on every SERP and product page, such as "customers who bought this item also bought," "save or upgrade with a similar product," "compare with similar items," and "frequently bought together."

"Related" features like these are especially useful for keyword research in highly competitive markets.

Search Operators

The site: operator is one of the most used operators for researching information on a site. It's a fast and easy way to see what pages Google has indexed for a website. Note, however, that Google normally limits the number of pages it will show. Google changes this limit from time to time and limits it has used are either 300 or 400 pages. An example of these results for mit.edu is shown in Figure 3-3.
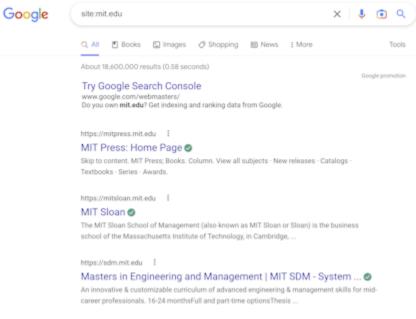

Figure 3-3. Sample site: query results

The pages shown by Google are usually the ones that are seen by Google as having more important pages. However, note that as Google does limit the number of pages shown so the absence of pages in this list does not mean that they are of poor quality.

The before: and after: query operators limit search results to those that were added to the index before or after a given date (in YYYY, YYYY-MM, or YYYY-MM-DD format). If you only specify a year, the rest of the date is assumed to be 01-01 (January 1st). You can use both operators in the same

query to limit results to a specific window of time. For instance if you wanted to see what people were saying about the Blizzard Entertainment game *Overwatch* between its official release date (October 4, 2022) and its launch (April 24, 2023), you could use this query:

```
Blizzard overwatch 2 before:2023-04-24 after:2022-10-04
```

You can also accomplish this via the Tools menu on the Google SERP (by selecting Any time, then selecting Custom range). That's fine for a single query, but if you're going to look up several keywords or use several different dates, it's quicker to use the search operators instead.

SEO Platforms

These services offer incredibly valuable tools for a wide array of SEO projects. Though the feature set varies among different platforms, the best ones can analyze your site for optimization gaps, suggest topics for new content, compare your site against your competitors, and find backlinks. Perhaps more importantly, they also harvest search data from Google, cleanse it of bot traffic, and connect it to in-depth keyword exploration features that can help you calculate the best opportunities.

SEO platforms are not free. Expect to pay anywhere between $100 and $500 per month for the good ones, depending on the level of service you need. It's usually free or inexpensive to do a limited trial, though, and the customer support is excellent (for the platforms we use and recommend, anyway). If search traffic is important to you, then this is money well spent. A professional SEO may even subscribe to several of them because they each do something a little better (or just differently in a way that you prefer) than the others, and certain projects may benefit from a particular niche tool or feature.

In order to do a proper valuation of your keyword list in Chapter 5, you must have a reliable and current data source for (at very least) monthly search traffic and CPC, though ideally you'd have keyword difficulty and current rank data from an SEO platform as well. We recommend any of the

services listed below for this purpose (and more – they're much more than just keyword research tools!), but there are several others on the market that might work better for your industry or region, and new ones may appear after this book has gone to press.

Semrush

https://www.semrush.com

Semrush is more than an SEO platform; it also provides tools for market research, social media management, content creation, and ad optimization. Within its SEO package, there are tools for keyword research, page optimization, local search optimization, rank tracking, backlink building, and competitive analysis. All of these services will be helpful for your SEO work, but we really want to highlight Semrush's superior keyword research capabilities.

If you only pay for one keyword research tool, Semrush should be a top candidate. Not that the others are bad (some are actually better for other purposes), but Semrush's biggest strength is its keyword database, which is the largest and most comprehensive in the industry, at more than 22 billion keywords (as of this printing). The database is built by scraping Google SERPS for the top 500 million most popular keywords, then analyzing the sites in the top 100 positions for each. The only potential shortcoming to this huge database is that it's only updated about once per month. That's fine for most SEO purposes (except for researching emerging trends), and it's on par with the update cycles of most other SEO platforms.

Semrush has a variety of keyword research tools:

Keyword Overview

A comprehensive report showing a keyword's popularity, difficulty, CPC, trends, and related keywords and questions. You can analyze up to 1000 keywords at a time by pasting them from your spreadsheet or text file into the Bulk Analysis field.

Organic Research

A report on the keywords used by the top 100 sites that relate to or compete with yours.

Keyword Magic

Based on the topic or keyword you give it, this tool will find every related relevant keyword in the Semrush database. You can also specify broad match, phrase match, or exact match modes to tune how many keywords that it provides.

Keyword Gap

Analyzes your competitor sites and identifies the best opportunities for keyword targeting and page optimization.

Keyword Manager

A real time analysis of how up to 1000 of your keywords perform on SERPs and with competitors.

Organic Traffic Insights

In combination with Google Analytics and Google Search Console, this tool can fill in the blanks for "not provided" keywords in your site traffic data.

The other Semrush features either speak for themselves or should be evaluated on an individual basis. In general this platform is well-documented and easy to use, but Semrush's customer support is among the best in the industry if you end up needing help.

Ahrefs

https://www.ahrefs.com

Ahrefs includes clickstream data from a wider variety of sources beyond Google, including Yandex, Baidu, Amazon, and Bing. Ahrefs doesn't offer

the big-picture marketing services beyond SEO like Semrush does, but its SEO toolset is feature-equivalent to Semrush:

Site Audit

> A comprehensive SEO analysis of your site, using more than 100 optimization criteria. The report shows technical issues with HTML, CSS, and JavaScript; inbound and outbound link impact; performance; and content quality analysis.

Site Explorer

> A report on your competitors' sites, including the keywords they rank for, their ad campaigns, and backlink analysis.

Keywords Explorer

> Enter up to 10,000 keywords to get search volume and other valuable data.

Content Explorer

> You provide a topic, and Content Explorer shows you an analysis of the top-performing articles and social media posts related to it. Excellent for finding backlink opportunities.

Rank Tracker

> Monitors your site's performance relative to you competitors. You can have updated reports delivered to you via email every week.

One nice feature in Keywords Explorer it has a large number of filtering options including a "topic" column.

Searchmetrics

https://www.searchmetrics.com

As with Semrush, Searchmetrics is a larger marketing services and consulting company that sells a powerful SEO suite. There are four aspects to the Searchmetrics Suite:

Research Cloud

> A domain-level market research tool that identifies valuable topics and keywords, analyzes competing sites, and reveals content gaps.

Content Experience

> Provides data that helps write effective search-optimized content for a given topic, including the best keywords to use, seasonal considerations, searcher intent, and competitive analysis.

Search Experience

> Performance monitoring and gap analysis for your site and your competitors' sites in organic search. Whereas Research Cloud is centered on your site, Search Experience is centered on the search journey that leads to it. This service has a much wider global reach than most others in this space.

Site Experience

> A technical audit of your site that reveals potential problems with search indexing, broken links, orphaned pages, and mobile responsiveness.

Searchmetrics Content Experience is outstanding for topic research. One feature that really stands out is the Topic Explorer. You provide one topic, and Topic Explorer creates a report containing search data and statistics, and a color-coded interactive mind map that shows how it performs relative to other semantically-related topics in terms of search volume, rank, seasonality, search intent, sales funnel, and level of competitiveness. You can drill down into any of those topics to get a more refined view of the keywords within them.

Searchmetrics is also among the few resources that can reconstruct "not provided" keywords in Google Analytics.

Moz Pro

https://ww.moz.com/products/pro

Moz Pro offers SEO tools for keyword research, rank tracking, site auditing, on-page optimization, and backlink building. Its main advantages are the Page Authority and Domain Authority scores, which help you find optimal backlink opportunities.

For keyword research, the Moz Keyword Explorer is an excellent resource, especially for natural language questions. It also has one of the best keyword difficulty scoring systems in the industry. The backlink research tool, Link Explorer, states that it has data on over 47 trillion backlinks.

Moz Pro also provides a site auditing tool that can be setup to perform a full site audit and identify a wide range of issues that may be hampering your SEO. In addition, you can setup alerts which will check your site and proactively let you know when problems are discovered.

Moz also offers a Page Optimization Score to help identify issues with your pages. This offers content optimization suggestions that enable you to improve the ability of your pages to rank.

RankRanger

https://www.rankranger.com

This is another marketing services suite that includes excellent SEO tools for keyword research, rank tracking, site auditing, etc. Its biggest selling point is its high degree of customization. Most SEO platforms have a few static presets for charts and graphs; RankRanger enables you to build your own. What we want to highlight in particular, though, is the superior natural language question research capabilities of the RankRanger Keyword Finder. If mobile search is a higher priority for your site than desktop search, and

you only want to pay for one SEO platform, RankRanger should be the first one you evaluate.

Other Platforms

As an SEO consultant, you may be asked to use the keyword research functions that are included in a marketing services suite that your client has already paid for, or is already familiar with. Here are some other good SEO platforms we've worked with in this capacity that we want to mention:

- Conductor Searchlight: https://www.conductor.com

- BrightEdge: https://www.brightedge.com

- SEO Clarity: https://www.seoclarity.net

Again – they do most of the same things, and you'll find that you prefer one over the others for specific purposes. There are many more SEO platforms and keyword research services than the ones we've listed in this book. These are just the ones we've used successfully and are comfortable recommending. If you want to use a platform that isn't covered in this book, some important questions to ask before subscribing to it are:

- Where is the data coming from, and how often is it updated?

- Does the data apply to the region or locale that my site is being marketed to?

- Can I import my whole keyword list, either via copy-and-paste, or by uploading a CSV?

- Can I export keyword data to a CSV (or XLS) file?

- Does it offer these metrics: monthly search volume, keyword difficulty (or keyword competition), CPC (cost per click), rank?

- What are its unique reporting capabilities?

KeywordTool.io

https://keywordtool.io

This service seeks to fill in the data gaps in the Google Ads Keyword Planner, which makes it a good secondary resource for optimizing Google Ads campaigns. Because it incorporates data from Google's autocomplete feature, KeywordTool.io is also particularly good for local keyword research.

Serp API

https://serpapi.com

This is a programmatic interface to Google and other search engines that enables you to run automated queries and returns SERP data in JSON format. You can plug that data into a dashboard or reporting engine or convert it to CSV or XLS and work with it in a spreadsheet. This is similar to what most SEO platforms do to scrape search data, except Serp API offers many more datapoints, customization options, and access to a wider array of search engines:

- Google
- Baidu
- Bing
- Yahoo!
- Yandex
- eBay
- YouTube
- Walmart
- The Home Depot
- LinkedIn

If you only need access to search data, and you have a Web developer or Python guru available to help, then Serp API is a cheaper alternative to a more comprehensive SEO platform.

Automation

Typically, retrieving keyword rank data from an SEO platform is a manual process: you provide a list of keywords, then filter and sort the data, then export it to a CSV or XLS file. From there, you'd import or copy/paste the data into a spreadsheet or analytics engine.

Some Google Sheets extensions exist that pull data from Google services like Google Search Console and Google Analytics. One of our favorites is Search Analytics for Sheets (https://searchanalyticsforsheets.com).

Some SEO platforms also offer access to their data via an API (application programming interface), which enables you to script part or all of the export process. If your preferred platform or data provider has an API, expect to pay extra for an API key and a certain number of monthly API usage units.

Some services allow you to pull just about any data from them via API – not just keyword reports. You can use this data in custom or third-party dashboards or business analytics packages, CMSs, or you can write a quick Python script to fetch your new keyword data CSV every month.

If you have a very large (by your own standards) keyword dataset, or if you maintain separate keyword lists for subdomains, content directories, or product lines, then you may find value in using AI to do the categorization and sorting for you. Specifically, we want to call out BigML (https://bigml.com) as an easy-to-use platform for building machine learning models for sorting large keyword datasets.

YouTube Optimization

YouTube is the world's second most popular website (Google being #1), and it has so much content that it's nearly impossible to navigate without its

built-in search function. Unfortunately – and unlike Google's video vertical search feature – that multitude of searches only leads to videos on YouTube, so its benefit to SEO is at best indirect (this is covered in more detail in chapters 10 and 11).

That said, if you're optimizing for YouTube search, vidIQ (https://vidiq.com) is one resource that can help. It's focused on gaining YouTube views and subscribers.

Another excellent YouTube resource to consider is TubeBuddy (https://www.tubebuddy.com). A couple of its most valuable features are the Keyword Explorer and A/B testing of video titles and thumbnails.

NOTE: Some SEO platforms (such as Ahrefs and RankRanger) also have YouTube keyword research and rank-tracking capabilities.

Conclusion

In this chapter we covered the SEO tools that we like to use (or in some cases are forced to use because of budget limitations or client preferences), but there are hundreds of other options out there that you may prefer for a variety of reasons. Perhaps there are tools that are better suited to non-English languages, or there are plugins for enterprise service or analytics suites that you'll have to learn to use because they're already deployed and paid for. You may also find that your preferred toolset for one project may not be as good a fit for a completely different project.

Beyond that, search technology, the Web in general, and the SEO field are all constantly evolving. New needs and concerns will arise as old ones fade away; new utilities and services will continue to enter the market, and some will merge with others or leave the market entirely. SEO platforms are very competitive with each other and will add new features and capabilities in the future, so even if you're an in-house SEO with only one site to manage, it's still a good idea to re-evaluate your SEO tools now and then.

Regardless of which platforms, services, and utilities you select, the only thing that matters is that you have the right tools to do SEO *your* way.

CHAPTER FOUR

SEO Planning

Planning your SEO strategy in advance is the most important step towards effective and successful SEO implementation. Planning is 90% learning and 10% predicting. Things never go perfectly according to plan, but if you are thorough in your process – if you put an appropriate level of effort into *planning* without relying too heavily on the accuracy of the *plan* – you'll be prepared to adapt to unexpected obstacles.

Your SEO strategy should generally address the following components:

1. Understanding the business you're working with and their objectives

2. Assessing web development, content, analytics, and management resources

3. Identify critical technical SEO issues

4. Identify the most efficient and highest impact improvements

5. Prioritize tactics and the phasing of implementation

6. Track, measure, and refine on a timeline that is supported by resources and overall business objectives

Developing an effective SEO strategy requires that you learn as much as you can about your client's (or employer's) business. In many cases, your initial plan will be limited to a few critical technical SEO tasks. After these fundamentals have been addressed, you'll be able to begin collecting and analyzing data that will guide you in developing a long-term plan for higher-level SEO tasks, such as competitive analysis, keyword research, seasonal planning, and content marketing.

The Business of SEO

Your SEO strategy is only as valuable as your ability to execute it. You can know exactly what's wrong with a site, and exactly how to fix it, but be financially or organizationally blocked from implementing the necessary solutions. Before diving into technical SEO implementation, it is crucial to address the important topic of business management.

> **NOTE**
>
> This section is only a basic overview of topics that are covered in much more detail in books about project management and business administration. For a more in-depth perspective on soft skills and client strategies for SEOs, we recommend *Product-Led SEO*, by Eli Schwartz.

This is true whether you are working in-house, as an independent consultant, or within an agency. In general there isn't anything different about the SEO consultancy industry, as standard business fundamentals

apply – but there are a few extra considerations when it comes to evaluating clients:

Retain an attorney or law firm knowledgeable in digital marketing and e-commerce.

> Eventually you will have a prospective client who may be doing something illegal (or something that will probably become illegal when it's publicly exposed). If there are even slight hints about the legality of a site you're asked to work on, ask your lawyer for an opinion before you agree to take the client. Retaining counsel is especially important when dealing with international clients, from both a contract as well as content, perspective.

Customize your agreement payment terms to ensure you get paid for completed work.

> There are numerous ways to structure client agreements, and payment terms are no exception. Some companies charge a flat rate paid monthly, with or without an initial cost, others get paid based on specific projects, while others get paid in other ways tied to deliverables over a short or long term. Consider the inclusion of late fees for payments not received by their due date.

Consider confidentiality and non-disclosure agreements (NDA) as needed.

> Both you and your client have intellectual property you both likely want to protect. Mutual confidentiality and non-disclosure agreements can be an easy way to protect this valuable IP and create a sense of safety within the relationship.

Clearly identify performance terms.

> Consider what work will be done, by whom, by when, and in what manner, as well as what support is needed by the client to ensure performance. The right attorney can draft a contract template for you to use with your clients which they will be comfortable enforcing and defending, and which will set you up with the appropriate protection.

Extensively document your work.

> Break down projects into tasks, and keep track of hours and expenses (with receipts). Establish analytics and measurement requirements and/or protocols specific to the client's needs and the overall SEO strategy.

Ethical and Moral Considerations

Every site can benefit from good SEO, even sites that contain and promote content which many large corporations consider to be of a harmful, terrorist, and/or violent extremist nature. These types of sites, and the entities behind them, are regularly identified by members of the the GIFCT (The Global Internet Forum to Counter Terrorism), including Facebook, Microsoft, Twitter, and YouTube), in their efforts to remove harmful content from internet circulation. There are many counter-efforts utilized (including SEO efforts) by those seeking to reduce the visibility of these content publishers.

Familiarize yourself with the type of content your client, or prospective client, is asking you to promote online to ensure that the client's content is in accordance with the various community and safety standards and guidelines on the platforms on which you choose to promote them.

The Escape Clause

As in all business relationships, despite your initial due diligence issues can still arise after the contract is signed and/or the project is funded. Some examples include:

- A change in management or ownership.

- The information you were given was incorrect (such as a deadline, sales target, budget, or asset / resource control).

- A critical internal partner refuses to participate (the Engineering Director or Product Manager refuses to assign designer or

developer time to SEO tasks).

- Other company priorities usurp the internal resources that had initially been assigned to SEO.

Often, these issues stem from internal client misconceptions about the complexity and importance of SEO, so make sure your requirements for client support are clearly articulated before the beginning of work, and as they are identified throughout the course of the engagement.

Typical Client Scenarios

Eventually every company with a Web presence knows that it needs to have greater search visibility. The SEO solution depends on how big the company is, how long it's been on the Web, how much damage has been done by SEO-unfriendly infrastructure, development and implementation, and how well they have understood the scope of the opportunity from a content perspective and created high quality content to meet that demand.

Startups (Unlaunched)

An unlaunched startup is as much of a "green field" as you'll ever find in the SEO industry. Since you are starting from scratch, you will be in the enviable position of having influence over various technology and development decisions which will directly influence the site's SEO success for years to come.

If you're an employee it's likely that your involvement in SEO is on a part-time basis and you have many other duties at the same time. Whether you're an employee or in a consulting role your role is likely to be advisory as you guide the company towards avoiding search visibility and content development issues, while also guiding overall usability, design and development in the initial startup phase.

Startups (Launched)

For startups which already have a web presence but have little or no SEO visibility, the first question you'll usually ask yourself is: "Improve it or replace it?"

Generally speaking, many startups often have limited human resources. For example, the administrator for the website may be the same person who serves as the IT manager and IT support administrator (two entirely different roles!), and this person is often tasked with managing the SEO efforts that the marketing team decides to pursue. Conversely, in a larger organization, the management of SEO would likely be divided into several roles filled by several people.

In this scenario, the scope of your role as an SEO will be dependent upon the organization's ability to understand SEO implementation requirements, and the level of investment available for SEO.

In all situations, it is important to remember that in order to be effective (and for the organization to have a chance at SEO success), you must have strong influence over the website's technology, development, and design decisions, and you must have the ability to effectively advocate for the resources needed to execute on those decisions.

Established Small Businesses

In trying to describe a "small business" for this section what we really mean is "a private company that doesn't intend to go public." Startups are usually heavy on early investment, and aim to either go public or get bought out by a bigger corporation. Sometimes, though, there's enough profitability to keep going, but not enough growth to attract an advantageous buyout. And then there are businesses that start small and never have the intention of going public or selling out – most often family-owned and sole-proprietor businesses.

The good news about privately-held businesses is that they're forced to be profitable; the bad news is, they also tend to be cautious when it comes to spending. Private business owners tend to think in different timeframes than public business executives; year-end sales totals and monthly expenses are

usually more important than quarterly profits. As with startup organizations, if you're responsible for SEO then you likely have several other responsibilities at the same time. If you're consulting for them you can avoid sticker-shock the same way car salesmen do: by spreading out your work on a longer timeline to keep monthly costs low, Either way you will need to prioritize tasks that will most reliably show month-on-month progress.

Small business organizations can be a nightmare. They often want to do things as cheaply as possible and have unrealistic expectations on timelines. Frequently, someone at the company has already tried to improve search rankings based on information from old articles and blog posts, and ended up making things worse. Then the search traffic stops, and you'll get a panicked call or email from the owner, begging you to fix it. Small business sites – especially e-commerce and local service providers (including a surprising number of lawyers) – that have been around for a decade or longer are a great place to find black-hat SEO gimmicks. If there has been a sudden drop in search traffic you should begin by investigating if the drop is due to a Google penalty.

It isn't impossible to undo the damage caused by unethical SEO hacks (we cover this topic in Chapter 8), but you can't make any guarantees. At best, sites that have been manually removed from the search index will take a lot of time and effort to repair.

Large Corporations

Whereas private companies present autocratic challenges for SEOs, large public corporations often sabotage SEO efforts with bureaucracy and inter-departmental warfare. Corporate environments may offer more resources for SEOs (in the form of money and in-house talent), but only if you're willing and able to play the political game. You might have a blanket approval from the CEO to do whatever is necessary to improve search traffic, yet still find yourself stonewalled by product managers and marketing VPs.

> **NOTE**
>
> Public companies are rarely as top-down as they look from the outside. In fact, in many cases you're better off thinking of them as *bottom-up.* Successful corporate employees only seek the approval of their direct manager; this paradigm continues upward, level by level, all the way to the C-suite.

To be successful, you must be able to build a cross-functional and multidisciplinary SEO team that spans departments and has the explicit support of Director-level management in engineering, IT, and marketing. The ideal SEO team consists of a manager (who may have a manager, director or higher title), designer (and possibly other creatives), software engineer (or a full Web development team), systems administrator (or IT representative), business analyst, and systems analyst. Ideally these people identify as SEOs or have SEO experience, but if not, they should be competent enough to learn how to attune their skillset to requirements and sensibilities of this field.

> **NOTE**
>
> Nearly every department can help with SEO in some way, but the only ones that can effectively block SEO efforts are engineering, marketing (including design and product management), and IT.

MBOs, BHAGs, and Other Corporate Flora and Fauna

The key to success in a large corporation is to integrate with its process for defining and tracking goals. Every company's culture is a little different, but all corporations that have a history of success use some kind of paradigm for defining very large goals, and breaking them down into departmental objectives. In order to get budget and resource commitments, you must figure out how to attach SEO tasks to every level of that hierarchy.

At the top of the food chain is the big, hairy, audacious goal (*BHAG*; we will leave it to your imagination and your Google skills to figure out the other possible variants of this acronym). This is typically something that seems impossible right now, but could be achievable in several years if there is focused and sustained effort. Some classical examples of BHAGs are:

- Find a cure for breast cancer.

- Put a man on the moon, and bring him back to Earth safely.

- Free Tibet from Chinese occupation.

- Organize the world's information.

One level below the BHAG is the manage by objective paradigm (*MBO*) popularized by the legendary author and business process consultant Peter Drucker, which establishes major business objectives (also *MBOs*, depending on your client's corporate vernacular) that work toward a BHAG, but are on a shorter timeline (usually one fiscal year). Some MBO examples based on the previously-listed BHAGs might be:

- Identify genetic and environmental factors involved with breast cancer.

- Design a spaceship that can go from the Earth to the moon and back.

- Achieve Chinese Communist Party recognition of the Dalai Lama as the foremost spiritual leader of the Tibetan people.

- Create a computer program that can find and record all relevant Web pages and categorize them according to a curated metadata model.

Potentially below the MBO, you might have a statement of objective and key results (*OKR*). This is a much more specific goal that defines exactly what someone will do to achieve a measurable result. For instance:

- I will develop a DNA test that can identify genes associated with breast cancer.

- I will design a spaceship that can propel a human-habitable capsule beyond Earth's atmosphere.

- I will create a media campaign that convinces members of the Chinese Communist Party that True Communism is best achieved by detaching from the unproductive socialism of Tibetan thought.

- I will engineer an automated script called Googlebot which will scrape an HTML page's text content, then follow all of its outbound links.

Also below (or sometimes on the same level as) MBOs are specific, measurable, achievable, relevant, and time-bound (*SMART*) goals. Basically these are OKRs that include a specific timeframe. For instance:

- I will create a saliva test that can accurately detect both a BRCA1 and BRCA2 gene mutation before the next anniversary of Susan G. Komen's birthday.

- I will design a multi-stage rocket that is capable of achieving planetary escape velocity before my Russian competitors.

- I will identify all potential ideological barriers to Tibetan inclusion in the One China vision before the next meeting of the National Congress.

- I will map the World Wide Web as a graph data structure before I get evicted from the Stanford dormitory.

Regardless of the methodology and acronyms, everything you do for your client's company should align with the hierarchy of goals that they claim to honor. At some point you will be asked to defend your SEO budget; the best way to do this is to show how every dollar is spent in service to the goals represented by the appropriate corporate acronyms.

Initial Triage

Congratulations, you got approval for the project (if you're in-house), or the contract (if you're a consultant) and you're ready to get started. Here are your action items for phase one, regardless of the type or size of company you're working with:

1. Figure out what's already been done in terms of SEO. Who was here before, what did they do, and if they're not here anymore why did they leave? In addition, map out how the company sees the role of content in SEO and they;re willingness to invest in that content going forward.

2. Identify the IT and SEO products and services that are in use right now, and determine whether you'll be locked into certain vendor agreements for your toolset. Learn what the content creation budget is and how flexible it might be going forward.

3. Implement some baseline analytics (or configure the existing analytics service, if there is one), and start collecting data.

4. Look for major technical SEO problems, then fix them. Identify content gaps and get the content created to fill them.

Document Previous SEO Work

If there are another SEO's fingerprints on this website, stop and take stock of what they've done. Ideally they did everything right and documented their work in a detailed log, which you can compare against analytics data to gauge the impact of each change.

One important aspect of investigating prior SEO work is determining what was done in terms of link-building activities. Were shady practices used that carry a lot of risk? Was there a particular link-building tactic that worked quite well? Analyzing the history of link-building efforts can yield tons of information that you can use to determine your next steps.

If no such log exists, start creating one yourself. Begin by asking your client if they have a copy of the previous SEO's recommendations, requests, or statements of work. Next, look for prior snapshots of the site on the Internet Archive's Wayback Machine (http://archive.org/web), and make note of what changed and approximately when.

NOTE

From this point forward, document every change you make, including all relevant details such as URLs, timestamps, and extra procedures such as server restarts and software upgrades. This may seem tedious in the beginning, but you'll thank us later when something goes wrong and you don't know which of the past 50 changes to revert.

Look For Black-Hat SEO Efforts

Occasionally you're going to find some amount of unethical SEO activity, otherwise known as *black-hat* SEO. The most common black-hat tactics are:

- *Keyword stuffing:* over-using specific keywords in your content to such an obnoxious degree that no human would want to read it. This term can also refer to dumping your keyword list into some part of a page (or, not uncommonly, on every page on a site), sometimes at the bottom or in invisible text.

- *Buying or selling links* without qualifying them properly as *nofollow*.

- *Cloaking:* showing Googlebot a search-friendly page, but redirecting real people to spammy pages. Sometimes this is referred to as a *doorway page* or *gateway page*.

- *Content theft:* stealing content from other sites, then either republishing it as-is, or spinning it (using a script to change a few words here and there to try to defeat duplicate content detection algorithms).

- *Thin content:* similar to keyword stuffing, this refers to useless content designed to appeal to search engines, but is worthless to humans. Often this is the same block of text published on several pages on a site, with the keywords swapped out for each page.

- *Spamming* links to your pages in the comment sections of blogs and on forums.

- *Using a black-hat script or service* to automate any of the above, especially content theft (scraping and spinning), thin content generation, content submission to sites that accept user contributions, forum and blog comment spamming, and auto-clicking on SERPs and ads.

These are just the old standbys; new dirty tricks are constantly being discovered and exploited. Unfortunately, black-hat tactics often work very well at first; when Google detects them, though (and it always does), it removes the offending pages from the search index and bans the site. *It's never worth it.*

The quickest way to find the most obvious black-hat SEO tactics is to use a backlink analysis tool such as Majestic SEO or Open Site Explorer to find links from irrelevant and/or low-quality sites. For instance, here are examples of two devious link schemes:

Black-Hat Case Studies

GiftCertificates.com's short-lived nemesis was FindGiftCards.com, which came out of nowhere to command the top two spots in Google for the all-important keyword *gift certificates,* thus relegating GiftCertificates.com to the third position. How did FindGiftCards.com do it? It operated a sister site on the domain 123counters.com, which hosted a free hit counter script

that propagated link spam across thousands of sites, all linking back to FindGiftCards.com and other sites in its network.

Sadly for FindGiftCards.com, co-author Stephan Spencer discussed the company's tactics in an article he wrote for Multichannel Merchant, and Google became aware of the scam. After that, all but two of its pages were removed from the index.

CraigPadoa.com was an affiliate seller that outranked SharperImage.com for its most popular product, the Ionic Breeze, by using frameset trickery and guestbook spamming (in other words: defacing other people's websites with fake guestbook entries that contained spammy links back to its own site). As soon as The Sharper Image realized what was happening, it jumped on the wayward affiliate. It also restricted such practices in its affiliate agreement and stepped up its monitoring of spam practices.

Site Changes That Can Affect SEO

Ideally your log should track *all* changes to the website, not just those that were made with SEO in mind, though you may not always have access to that information, especially if IT and engineering services are outsourced.

In larger organizations many different people can make many changes to the website that can impact SEO. In some cases they don't think the changes will impact SEO and in other cases they don't think about SEO at all. Here are some examples of basic changes that can interfere with your SEO project:

Adding content areas/features/options to the site.

> This could be anything from a new blog to a new categorization system.

Changing the domain name.

> If not managed properly with 301 redirects, this will have a significant impact.

Modifying URL structures

Including the Web server directory structure

Implementing a new content management system (CMS).

This will have a very big impact on SEO. If you must change your CMS, make sure you do a thorough analysis of the SEO shortcomings of the new system versus the old one, and make sure you track the timing and the impact so that you can identify critical problems quickly.

Establishing new partnerships

Relationships that either send links or require them (meaning your site is earning new links or linking out to new places).

Making changes to navigation/menu systems

Moving links around on pages, creating new link systems, etc.

Content changes

This includes publishing new content, revising existing content or deleting old content. These all can have a significant impact on SEO.

There probably isn't much you can do about changes that happened more than a few months ago, except perhaps redirect broken incoming links to new URLs and begin the work of rebuilding the lost SEO traffic. More importantly, we mention these common site changes because you will now have to be vigilant for them. You aren't necessarily going to be notified (let alone consulted!) about changes that other people in the organization (or its outsourced partners) make.

Identify Technical Problems

A large part of SEO is a technical process, and as such, it impacts major technology choices. For example, a CMS can either facilitate or undermine your SEO strategy. The technology choices you make at the outset of developing your site can have a major impact on your SEO results.

Servers and Hosting

Whether you host your own server on-site, co-locate it in a managed facility, or rent a server from a hosting provider, make sure it's built to handle the level of traffic that you expect to have. In many instances, it may already be underpowered for the traffic you already have.

Googlebot makes note of page load times, and while it may wait around a little while to see what's on a slow-loading page, a human often will not; this is what's known in the industry as "a poor search experience." If your site loads slowly, this can have an impact on how you rank in the search results (learn more about this in Page Experience in Chapter 6).

IP Addresses

Shared IP addresses (which are often a side-effect of shared hosting services) are not necessarily bad, but they're at very least sub-optimal. It doesn't cost very much to get a static IP address, even if you have to change your hosting provider.

First, though, check to see if that IP address (or the one you're currently using) used to belong to a blacklisted site (you can do this by using a service such as https://whatismyipaddress.com/blacklist-check). As search engines crawl the Web, they frequently find entire blocks of IP addresses filled with nothing but malware and spam. In these instances, the entire IP range may be excluded.

Bandwidth Limits

Hosting providers limit the amount of incoming connections and outgoing data transfer (generally referred to collectively as *bandwidth*) for each server or account. You can usually purchase a higher service level to raise your bandwidth cap; if not, then you'll have to find a better hosting solution. If you don't have enough bandwidth, and you have some piece of content that suddenly goes viral, your site will either load very slowly or not at all. If you're lucky, it'll go back to normal when people stop trying to see your content; if you're not lucky, your site will stay offline until your monthly bandwidth allocation is replenished.

> **NOTE**
>
> Be wary of hosting providers that claim to offer "unmetered" bandwidth. This doesn't usually mean what you think it means. Most often it means that you still have a (sometimes secret) bandwidth limit, but the provider won't keep strict track of how much you've used. It may also mean that bandwidth is throttled down after a certain threshold, which will cause your pages to load slowly.

Gated Content

Making content accessible only after the user has completed a form (such as a login) or made a selection from an improperly implemented pull-down list is a great way to hide content from search engines. Do not use these techniques unless you want to hide your content from the whole Web. Also, do not attempt to use cloaking techniques to show the content to search crawlers, but not to human visitors.

If your business model is to get people to pay for content then make sure that a significant amount of content is visible without going through the paywall in order to have a chance of still earning some organic search traffic. In addition, you can use flexible sampling as a way to let Google crawl and index your paywalled content. This does require that you let users who see that content in the Google results are able to read the entire piece without the paywall but you can limit how many times they get to do that per month. You can read more about this in Chapter 6.

Temporary URL Redirects

When you change a page URL, you have to instruct your web server to resolve an existing URL to a new one, via a redirect. Redirects can be server-side, or done via meta refresh (with the former generally the preferred method unless unsupported by your web platform). If your URL change is a temporary one, then you would use a temporary redirect command in your web server configuration (usually a 302, 303, or 307 type server-side redirect).

Most often, however, URL changes are permanent. The correct HTTP code for that is a 301 or a 308 "moved permanently" redirect. When search engines find a "moved permanently" redirect, they view this as a "strong" signal that the redirect target should be canonical and update the index with the new URL, as well as pass the majority of the link authority from the old URL to the new one. If they find a "temporary" redirect, they interpret this as a "weak" signal that the redirect target should be canonical, they do not update the index, and the old link does not pass its link authority to the new one.

Some web servers, content management systems, and web-based publishing platforms use 302 redirects as the default for URL redirection - if so, you should change this to 301 immediately.

Mobile Responsiveness

The majority of internet searches come from a mobile device, and Google no longer maintains separate indexes or crawlers for desktop and mobile sites. At minimum, your site must be mobile-friendly in order to succeed.

If you're building and launching a new web site, it should be designed to be mobile-first, optimizing for speed and mobile usability across different device types, screen sizes, and mobile operating systems.

For existing web sites which haven't been substantially redesigned in the past two years, it is recommended to set aside time for an audit to assess your current search visibility, and determine the site's mobile-friendliness.

Know Your Client

Now your SEO planning shifts from the objective to the subjective. Here are some useful questions to help guide you in the right direction, optimally considered from a non-technical frame of reference:

What are the company's current objectives, and what are its future goals?

> This might align with the official mission statement, and it might not. For instance: Red Bull is manufacturing and selling energy drinks, and

hopes to expand its market share. A pre-launch startup is developing a product that it hopes to bring to market on a specific release date.

What are all of the identifiable outcomes that the company's success relies on?

For instance: a closed deal, product sale, mailing list subscription, social media follow, download, or service sign-up.

Who are the people (your users) who are likely to perform those ultimate actions?

Who is this company trying to attract or influence? Who are its customers, clients, users, or followers? What are their demographics? Where do they live? What languages do they speak?

What are all of the current and possible methods of connecting people to those ultimate actions?

Are there online and offline advertising campaigns, discount sales, seasonal opportunities, bulk mailings?

How does the company track and measure success?

There can be several answers to this, but there should be one metric that takes precedence. Gross revenue or return on investment (ROI) are usually factored into this measurement, but within the context of revenue generation there are various identifiable user actions which, when tracked and measured, can generally correlate with broader financial success over various timeframes.

What does the competitive landscape look like?

Who are the market leaders in your market space? What are they doing that you're not? What are the search terms that you're winning on and what can you learn from that to guide your SEO strategy going forward.

Take Inventory of the Client's Relevant Assets

Determine the current assets available for SEO is the first place to start when planning your overall strategy. Even if your client's business is relatively small or new, there is likely still a cache of valuable material that you can use for planning and keyword research, such as:

- Lists of products, brands, and trademarks

- Commissioned studies and paid market research reports

- Customer interviews, reviews, testimonials, and surveys

- Licensed content (photos, videos, music)

- Analysis reports of previous marketing campaigns

- Access to Google Analytics (or other marketing/analytics platform) data

- Access to Google Ads data

- Site architecture or information architecture documents

Other assets that require more consideration are listed in the subsections below.

Customer Personas

A customer persona is the archetype of your ideal customers or clients. In some industries, this may be referred to as a buyer persona or a customer profile, and your company may have already defined them. If so, use the ones that you already have.

If you need to create your own personas, then consider these questions of your ideal customers:

- What is the age range for each persona?

- Where do they live?

- What is their job title (or role within the organization you're selling to)?

- What are their hobbies?

- How much money do they make (or what is their net worth range)?

- Are they married or single?

- Do they have kids (and if so, how many)?

- Do they primarily search from a mobile device, or a computer?

- How much money do they spend annually on your industry's products and services?

Next, consider some broader, more subjective questions:

- How late do they work?

- What do they do on the weekends?

The more details you can think of, the better. Then give this persona a one- or two-syllable name that is easy to remember, and record all of this information in a text file. Create a persona to represent each different type of ideal customer.

Enumerating all of this information will help you divine each persona's pain points, and that will help you understand the search queries they'll use and the questions they'll ask.

Domain Names and Subdomains

If you have multiple domains, some relevant considerations for most SEO projects would be:

- Can you 301-redirect some of those domains back to your primary domain, or to a subfolder on the site for additional benefit?

- Be sure to check the domain health in Google Search Console and Bing Webmaster Tools before performing any 301 redirects, as you want to ensure the domain in question has no penalties before closely associating it with your main domain(s).

- If you're maintaining those domains as separate sites, are you linking between them intelligently, and appropriately?

If any of those avenues produce potentially valuable strategies, pursue them—remember that it is often far easier to optimize what you're already doing than to develop entirely new strategies, content, and processes.

Vertical Content

Take the time to properly catalogue all of your digital media files. From an SEO perspective, it's useful to think of digital media such as images and video as *vertical content* that is separately indexed. A YouTube link is a URL that points to a page, from a search indexation perspective, it's also a video – especially when considering that the same video hosted on that URL is viewable within the YouTube mobile app. Regardless, you cannot make changes to the HTML or any of the code on a hosted service like YouTube, you can only optimize the content you upload and the metadata for your uploaded files.

When you control the hosting, however, vertical content can be a secret weapon for SEO. There may be high-difficulty keywords that you don't have the budget to compete for in Web search, but which offer affordable opportunities in verticals like images or video. Consider the Chapter 2 example of a query that has vertical intent: *diamond and emerald engagement ring.* The product vertical results are mostly huge brand names (Diamondere, Tiffany & Co.) who've undoubtedly paid a lot of money for top placements. As we explained earlier, though, this query suggests a search for both products and images. If the product vertical is out of reach, then you can more affordably target the images vertical.

Offline and Non-Indexed Content

You may have articles, blog posts, and videos that were never published on the Web – or even never published anywhere. Even if content was published on the Web, if it was never indexed by a search engine then it effectively was never published.

In the copy room filing cabinet there could be hundreds of well-written articles that were only published in a print newsletter which are still relevant enough to be reused. For example, you could publish the entire cache of newsletters in a searchable archive on your website. The same concept applies to email newsletters that were never published in a searchable archive.

Competitive Analysis

Once you have performed in-depth keyword research (or can rely on the research done before you), you are well positioned to begin performing some competitive research. Business owners usually have a very good idea of who their direct competitors are (regardless of search result rankings), so this can be a useful starting point – especially for companies that serve specific regions, locales, or niche markets.

In a larger company, the marketing and sales departments should be an excellent resource on this topic. Marketing, at least, should have some competitive analysis reports or presentations that you can learn from.

> **NOTE**
>
> High-ranking competitors aren't necessarily good examples of SEO, so be careful where you draw your inspiration from. Don't make any big decisions or assumptions about competition until you've seen good keyword data.

Information Architecture

All of the pages and digital media resources that comprise your website must use concise but descriptive filenames, contain appropriate metadata, and be linked to from other pages on the site. From a usability perspective, users on any device type should be able to quickly and intuitively navigate to your homepage and to the information they're looking for. If a user can find your content easily, so can a Web crawler. That's because Googlebot discovers new pages by following links and analyzing page content.

Information architecture encapsulates everything described in the previous paragraph: nomenclature, taxonomy, and navigation for a large collection of information. It's a subset of library science, and it's useful to think of your website in similar terms. If webpages and vertical digital media assets were books, and your website were a library, how would you organize it such that visitors do not need an index to find a specific resource?

NOTE

If you're having difficulty with this, you may need to engage with a Systems Analyst.

For planning purposes, you must create a site architecture document that defines or proposes a file naming convention, and illustrates the directory structure and content taxonomy.

Advanced IA Development

A good taxonomy has as few categories as possible, without defining them so broadly that the classification becomes useless. You will be in a much better position to flesh out your site architecture document after you've gone through the keyword research process (covered in Chapter 5).

Once you have your site architecture mapped out, here are some refinements and addendums to consider:

Cross-references should be cross-linked

> Many of your pages and assets can reasonably fall into more than one category. Beyond that, keyword research will yield many associations that you wouldn't be able to anticipate on your own. For example, look at any product page on Amazon.com and note how many ways products are cross-referenced (frequently bought together, customers who bought this item also bought, etc.). Hashtags are another form of cross-referencing that you should consider.

Descriptive anchor text

For all internal links, avoid using irrelevant anchor text such as "More" or "Click here." Try to be as specific and contextually relevant as possible and include phrases when appropriate within your link text.

Implement breadcrumb navigation

This is the best way to show users where they are in the topical hierarchy, and an excellent way to properly use keywords in the anchor text for each category page.

Refactor your architecture to minimize link depth

The closer a page is to the base URL (the domain name), the easier it should be to navigate to from the homepage. From a human (and search crawler) perspective, the most important things are always close at hand (your driver's license, house key, and smartphone, for instance), and the least important things are archived or stored in unobtrusive places (receipts from nine years ago, the owner's manual for your car, out-of-season holiday decorations). Therefore the deeper a URL is in the site hierarchy, the less important it is assumed to be from a search perspective.

NOTE

You may have to make some compromises in good information architecture in order to reduce link depth for SEO purposes. Some sites will need to be as flat as possible in order to minimize the number of clicks from the homepage.

Measuring Progress

Websites are complicated in their design and development, and there are many hidden factors among the services that enable them. Search engines are quick to crawl a site and discover new content (assuming the site is search-friendly with appropriate XML sitemaps submitted via Google Search Console), so it is imperative to create properly configured sitemaps

and keep them updated. You usually won't know exactly how, when, or if a particular change will impact search visibility and keyword rankings.

There are various elements to measuring SEO progress and demonstrating the value of investing in SEO: appropriate planning, strategic and efficient implementation,documenting changes, and tracking and measuring relevant data. You cannot afford to let the organization take it on faith that you're doing the right things, especially when there aren't often immediate results in search.

Technical SEO improvements tend to have quicker results than on-page SEO enhancements because they remediate fundamental problems with crawling and indexing. Think of a site with technical SEO problems as a race car with two bad spark plugs; replacing the failed components will of course lead to instant performance gains, but in the larger context of a race, all you've done is return to the baseline level of performance. Winning the race will take a lot more time and effort than simple maintenance tasks.

Much of your ability to successfully measure the impact of your SEO efforts will depend on the analytics platform data you have access to, so take charge of website analytics, and leverage any other marketing/performance analytics platforms if you can. Decision-makers and internal influencers at a company can sabotage you with junk metrics like "hits" or "mentions" or other things that aren't relevant to the company's actual success criteria. You should be able to show, at any given time, how what you've done has led to, or will lead to, progress toward the company's goals.

Conclusion

The perfect plan can only be created in retrospect, after the job is done and every problem is solved. It is therefore imperative to learn as much as you can at the outset about your client's web site, target market, business objectives, and organizational structure. Further, from the moment you begin, document the relevant portions of your work and set up the appropriate analytics so you can track and measure your progress.

CHAPTER FIVE

Keyword Research

The Words and Phrases That Define Your Business

Keyword research is the process of finding the words and phrases that connect your customers or clients to your business. This is the most important aspect of search engine marketing, but it also has a great deal of business value beyond the Web. Keyword research enables you to predict shifts in demand; respond to changing market conditions; and provide new products, services, and content that Web searchers are actively seeking. In the history of marketing, there has never been a more efficient and effective

method of understanding the motivations of consumers in virtually every niche.

In this chapter, we'll walk you through the entire keyword research process, beginning with building your domain expertise and analyzing the language associated with your company and its products. Next we'll give you some tips for developing a list of topics, and how to use it to develop an initial keyword list. We'll show you how to set up a spreadsheet to track and calculate your keyword plan, and how to use data from an SEO platform to calculate the best opportunities. Lastly, we'll provide some guidance on how and when to update your keyword plan.

By the end of this chapter you will have a spreadsheet with search terms and phrases that are of value to your business, categorized by topic, and ordered by their level of opportunity. *This is the basis for every subsequent chapter in this book.* Without going through the keyword research process, you cannot develop meaningful plans for updating your website to be more search-friendly, or effectively create new content to attract search traffic, or successfully tune your analytics to look for future opportunities.

Don't rush through this. Keyword research is not a fast process, especially if you have a large retail site with a lot of different products. Plan to spend at least a few days on the initial effort.

NOTE: In order to do a proper valuation of your keyword list, you must have a reliable and current data source for (at the very least) monthly search traffic and CPC, though ideally you'd have keyword difficulty and current rank data from an SEO platform or data provider as well. If you haven't made a decision about that yet, then you can either use some of the free or low-cost tools we introduced you to in Chapter 3, or delay your keyword research effort until you're ready to sign up for an SEO platform.

The Different Phases of Keyword Research

The keyword research process can vary quite a bit depending on where you are in your SEO project, and the individual needs of the business. Therefore we're going to present the whole from-scratch process that you'd follow if

you were working on a completely new site that isn't currently ranking for anything. Even if this is not your situation, you should read this chapter in sequence anyway, just in case there's a knowledge gap.

The first keyword research effort is typically concurrent with an initial site audit that establishes a starting point for the big-picture SEO project, but every project and client is different. Ultimately it depends on the information you are required to deliver as part of your first estimate. This can include a basic assessment of the keywords that a site currently ranks for, but it should go beyond that to include a list of keywords that would likely benefit the company. If you are expected to produce a list of title or meta tag recommendations as part of your audit, then you'll have to put a lot of effort into your keyword plan.

After you complete your initial optimization work, you will be able to more accurately assess the cost of optimization for your keyword list. This is a good time to drill down into long-tail keywords to look for the most efficient opportunities. Depending on the size and complexity of the SEO project, this phase of keyword research can be the most labor-intensive.

Once you've built and refined your keyword plan, you'll schedule regular reviews to update it with new data. This can be done once a month or once a quarter; you should never go more than a year without a keyword review.

In most cases, seasonal keyword research should be tracked and scheduled separately from your regular keyword reviews. Nearly every site can benefit from seasonal keyword research, even if there isn't an obvious seasonality to the company.

Expanding Your Domain Expertise

Before you go further, ask yourself how well you know this company and the industry it's a part of. You cannot successfully conduct keyword research for a business that you aren't familiar with. You don't have to be the world's foremost expert on it, but you should reasonably understand the fundamental technologies behind the products and services it sells, its history, its mission, the size of the market, who the main competitors are,

the impact of various seasons and holidays, and basic customer demographics. If there are existing customer avatars or personas, ask to see them.

If you're an outside consultant, this is a critically important part of the process. If you're an employee of this company, then take the time to learn something new about it. Talk to the most recent hires and the longest-tenured employees about their experiences, read the documentation, do a Google News search and see what's been written about it.

Building Your Topics List

It's a good idea to build a topics list before you go further with keyword research. Start by asking yourself what the entities and concepts are for your products or services, your company, and the larger context of the industry.

If possible, begin at the lowest level, then go up one step at a time. What are the topics that apply to every search query that you want to rank highly for? For instance if you're selling smartphone cases and screen protectors, then the lowest level might be "smartphone accessories," or if that domain is too broad (if you don't sell earbuds, chargers, or cables) then perhaps "smartphone cases." Think of these root-level topics as *domains*.

Next, ask yourself what the related topics are. Why do people buy smartphone accessories? What are their concerns? What purposes do these products serve? What are the alternatives? What are the most important features or qualities?

As a solution provider you're used to solution-side thinking; you're at the end of the chain. Your customers started their journey long before this, though, so you want to be visible to them as early in the process as possible. In some instances this may be before they even know they have a need or a problem yet.

Let's say you sell mobile phone cases and screen protectors for Apple and Samsung phones. The problems that lead people to buy a case or a screen

protector are pretty obvious, but some of the alternative perspectives aren't. Your customer may be someone who's dropped an expensive mobile device and is searching for a repair option, but consciously or subconsciously they're also asking themselves how they can prevent this from happening again. That's a great time to show up in their search results.

A contextually similar but topically separate concern is water damage. If someone is searching for how to tell if their phone has been water-damaged, or how to properly dry out a phone that fell into a toilet, that's also a great time for your waterproof phone cases to show up in their search results. They aren't looking for a waterproof smartphone case right now, and in fact they may not even know that such things exist until they see your site come up in search results for *what kind of rice to use to dry out a soaked iphone*.

Insurance is an alternative path that someone might explore after suffering a smartphone loss, but it's expensive. You might position your products as a cheaper preventative option to a monthly insurance fee. The main question a customer would probably ask is: "Is smartphone insurance worth it?" You could get a lot of sales from ranking highly in those results even though the query is only tangentially related to the products you sell.

So your related topics might be:

- smartphone screen damage

- smartphone protection

- smartphone insurance

- waterproof case

- iPhone compatibility

- Samsung compatibility

- stylish case

These aren't keywords, they're classifications for keywords (though there may be some overlap between the two). Since keywords will eventually provide more context, you can simplify and generalize your topics by

assuming that they are subsets of your domain (which is "smartphone cases" in this example):

- screen damage

- protection

- insurance

- waterproof

- water damage

- iPhone

- Samsung

- style

Among these, are there any that would apply to a disproportionately large number of other topics (or in mathematical terms: topics with a *high cardinality*)? In our example, the two that stand out are Samsung and iPhone because (assuming you only sell cases for these two brands) one and only one of them will always apply to every other topic. When you discover these *superset* topics, make note of them and keep them separate from the main list. When you start building out your keyword spreadsheet, you'll create columns for each of them so that you can do fine-grained sorting and filtering.

With the realization that Samsung and iPhone together span 100% of your topics list but have no overlap, it makes sense to go one level lower to their common domain: *device brand*. Even if you have a few products that are brand-agnostic (such as a screen protector that is compatible with some iPhone and Samsung models), the taxonomy still requires one and only one phone brand per keyword because people only search for one or the other.

Now take another look at the topic list. Are there any other high-cardinality topics that you would want to drill down into and sort by? In our example, the answer is "no" except for one item: *style*. Everything else is a Boolean; a case is either waterproof or it isn't, and screen damage, water damage, insurance, and protection refer to searches peripheral to your products. *Style*, however, has several important subcategories: color, materials, thickness, hardness, special features. Among those, you could drill down even further. If you think that you'll need to sort your keyword list by any of those subtopics, then mark this as a superset. If you end up being wrong, no big deal – it's easy to add or remove spreadsheet columns later.

When the remaining topics have a similar cardinality, or if it doesn't make sense to break them down any further right now, then whatever's left is your general list of topics. This is as far as you need to drill down in this example. There's more work to do later, but at this point the foundation is solid.

If you're having trouble coming up with topics, refer to the "Researching Trends, Topics, and Seasonality" section later in this chapter.

Preparing Your Keyword Plan Spreadsheet

If your keyword list is not yet in a spreadsheet, then now is the time to migrate to one.

Start with a new file. Label your first worksheet as *Raw Keywords*, and put every keyword you've come up with into column A. This is only an initial, unrefined list of potential search terms.

A quick reminder: "keyword" is a general term for a group of related words that will be used in a search query. For example, here are four separate but similar keywords of varying lengths:

- pink waterproof iphone case

- waterproof iphone case

- pink iphone case

- iphone case

Many of your keywords will be as similar as those are. Even though they're almost the same, and may lead to largely the same set of results, they all have different search data associated with them, and will provide different levels of value and opportunity. In fact, you could simply change the word order and come up with four more keywords, all with different search data. For now, go with whatever makes sense to you. When in doubt, choose the keyword that most resembles a natural language question.

> **NOTE**
>
> This is not the final list; you only need it to "prime the pump," so-to-speak, so don't exhaust yourself trying to think of every possible search query. When you eventually load these raw keywords into your preferred SEO platform(s), you'll be able to see the traffic levels, rankings, and difficulties for them and their many variations, so there's no sense in trying to go further than this right now.

Next, create a new worksheet tab called *Keyword Plan*. This will be your main worksheet containing your refined list of keywords and their metadata. For now, all you need to do is set up the structure. At the top of this worksheet, create a 10x2 table with a header, and populate the header row with the following titles:

- Keyword

- Monthly search volume

- Priority

- Relevance

- Difficulty

- Rank

- CPC

- Superset

- Topic

- Persona

- URL

The *keyword* column will contain the refined list of worthwhile keywords that you'll eventually import from an SEO platform or search data provider.

<div style="border:1px solid #999; padding:1em;">

NOTE

Don't put your raw keywords here – put them in the Raw Keywords worksheet.

</div>

Monthly search volume is exactly what it sounds like: the average search volume for the previous 12 months. Many SEO platforms can also provide indications of seasonality, or seasonal monthly changes in search volume.

Relevance refers to a subjective assessment of how relevant this keyword is to your current page content. This is covered in more detail in the "Keyword Valuation" section later in this chapter.

Priority is either a flag or a rating that identifies topics and keywords that are of major importance to your company, usually in alignment with your company's major business objectives, marketing initiatives, and/or sales

targets. This is covered in more detail in the "Keyword Valuation" section later in this chapter.

Keyword difficulty (sometimes called *keyword competition* instead) refers to the relative amount of competition for this keyword. If a lot of sites are fighting over the same keyword, then it has a high degree of difficulty. It will take more work to place higher in the SERP for that keyword, and ads will cost significantly more. For instance, in the above example, the broad keyword *iphone case* will almost certainly have more competition (and therefore a higher keyword difficulty rating) than the more specific *pink waterproof iphone case*. There are a few different ways to represent keyword difficulty, depending on which data provider you use (that's covered later in this chapter), but the easiest standard to conform to is a scale from 1 to 100.

Rank refers to the SERP position that a page on your site (which is specified in the *URL* column) currently has for this keyword. These columns are not strictly required right now, but you'll need them later so that you can see your progress. Ranking data comes from SEO platforms that analyze your site. Other keyword research tools may only provide non-site-specific data for keyword popularity, difficulty, and CPC, which is all you need to develop an initial keyword plan. If your pages are mostly unranked (or ranked very low), or if you are launching a new site, then these columns will be empty anyway. Most data providers only include the top 100 pages for each keyword, so this column will usually be a range from 1 to 100.

CPC means "cost per click," which is the amount of money you'd pay for a clickthrough on an ad that targets this keyword. If you don't have access to good data for keyword difficulty, then you can generally use CPC as a substitute, though ideally you'd have both.

The *superset* column is a placeholder for a high-cardinality topic. In the previous example we defined two supersets: *device brand* and *style*. That means you'd rename this column to "device brand," and create a second column for "style." Add a column for each superset that you defined.

Topic refers to the refined list of topics that you created earlier. This sheet is for keywords, though, so in this context a topic is a *dimension* of a keyword – a category to which it belongs. That means that you have to list your topics somewhere else. Create a third worksheet tab labeled *Topics and Supersets*. In it, convert column A into a table with the name *topics_list*. Change the column title from Column1 to *All Topics*. Then go back to your *Keyword Plan* worksheet, select the *Topic* column, then use the *Data Validation* feature to allow only a list, and use this formula as a source:

```
=INDIRECT("topics_list[All Topics]")
```

This will avoid potential filtering problems by strictly enforcing the correct nomenclature in the Topic column. This makes filtering easier, and it also makes it possible to pivot the data table to aggregate all the keywords for each topic, which you may find useful later. If you want to add new topics in the future, you must add them to the All Topics column in the Topics and Supersets worksheet.

Repeat this process for each of your supersets (go to your Topics and Supersets worksheet, add a new column to the topics_list table for each of your supersets, retitle the columns to match each superset, populate each column with the items in those lists, and enable Data Validation for the corresponding columns in the Keyword Plan worksheet).

> **NOTE**
>
> If the data validation becomes burdensome at any point, disable it.

Persona refers to the customer avatars or personas you created or acquired from someone else at the company back in Chapter 4. You may find value in repeating the Data Validation process for the *Persona* column (and creating another worksheet tab for Personas), since that could be considered an abstraction of the *customers* domain. This is probably a column you would want to filter, sort, and pivot by. If you chose not to create customer avatars, then you can remove this column.

The *URL* column will contain the URL of the page that this keyword is currently ranking for. If a page is unranked, this will be blank. This column isn't required for developing an initial keyword plan, but it's useful later

Depending on the nature of your site and your industry, you may want to consider adding a column that identifies a keyword as being either *branded* or non-branded. This is just a flag for branded keywords – you can mark it with an *X* or a *1* or whatever you prefer. No search data will be imported into these cells.

You now have the basic spreadsheet framework to build out your keyword plan. You may want to make a copy of this file to use as a template for future projects. To the extent that it is relevant to your project, try to create the spreadsheet exactly as described, even if you think that you won't use some of this data. You don't truly know what you will and won't need in your Keyword Plan until the project is complete, and every project has different needs.

Internal Resources for Keyword Research

In Chapter 3 we covered a variety of third-party tools that can help you identify topics, keywords, and questions. Eventually this is where your keyword data will come from, but it works best when you provide a comprehensive list of raw keywords as a "seed." In addition to your own brainstorming, there are many potential keyword resources in-house.

It may be useful to make note of your sources. If so, go ahead and create a *source* column in your *Raw Keywords* worksheet to identify where or whom the idea for each keyword came from. It isn't critical that you keep track of this, but it could be helpful during future keyword reviews, or for other marketing purposes beyond SEO.

Gathering this kind of intelligence is what a traditional marketer might have done prior to initiating a marketing campaign before the Web existed. And of course, if any of this data is available to you from other departments of the company, be sure to incorporate it.

Web Logs and Analytics

Collect all internal Web traffic data you possibly can. If you can get a CSV export of useful data, do it, and add it as a new tab in your keyword spreadsheet.

The most valuable data from a Web analytics package is a list of search queries or keywords from incoming search traffic, but any page-level information about site traffic is also potentially useful, including visitor counts (daily, monthly, and seasonal), visitor location, visitor platform (browser and operating system), bounce rates, inbound link (or "referrer") URLs, direct link URLs to assets such as images or videos, and 404 "page not found" errors.

Competitive Analysis

Your competitors face the same challenge with keyword research and content optimization, and unless you are very lucky, they are probably also resourceful and creative. Even if they haven't invested in SEO, at the very least they've put a lot of effort into learning about their customers and the best ways to appeal to them.

Review your competitors' websites and try to determine the keywords and phrases they're targeting for the products and services that compete with yours. Look for unique variations and synonyms they use in their content. Do these unique terms indicate shifting trends in the vernacular of your industry? Are they obviously optimizing for certain keywords? What non-brand terms do they use for their business? Have they written any articles or blog posts? What does the media say about them?

People

Every employee could have valuable insight into the thoughts and actions of your customers. You'll find a lot of value in talking to them, not just to get ideas for keywords, but also to reveal gaps in the concepts and terminology used by your organization and your customers.

Some basic questions you might ask are:

1. What are the words and phrases that define our business and its products or services?

2. What words and phrases do customers use when they talk about our products or services?

3. What are the questions that prospects and customers ask us?

4. What are the questions that people ask before they connect with us?

You

Before you meet with anyone, consider your topics, then generate an initial list of terms and phrases that you think are relevant to your industry and what your site or business offers. Include all of your various past and present brand names, products, and services. If your site has a massive number of products, consider stepping back a level (or two) and listing the lower-level categories and subcategories.

Aim to come up with at least a hundred keywords and phrases that could potentially be used in a search query by relevant, qualified customers or visitors. Ideally you'll come up with a list (or a series of lists) that looks a lot like Bubba's lengthy enumeration of the many ways that "the fruit of the sea" can be prepared in the movie *Forrest Gump:*

- Barbecue shrimp
- Boiled shrimp
- Broiled shrimp
- Baked shrimp
- Sauteed shrimp
- Shrimp kabobs
- Shrimp Creole

- Shrimp gumbo

- Pan-fried shrimp

- Deep-fried shrimp

- Shrimp stir-fry

- Pineapple shrimp

- Lemon shrimp

- Shrimp and potatoes

(This assumes that *shrimp* is the common entity.)

For now, stick to broad two- or three-word phrases like these. You can drill down into one keyword and expand it with relevant peripheral words and disambiguations if you really want to, but generally it's best to do that later when you have access to search data. You wouldn't want to spend an hour thinking of 50 more keywords based on *lemon shrimp* if that isn't what your business wants to target, or if it ends up being a low-volume, low-opportunity topic.

Everyone

If the business is small enough that it's logistically possible to call an all-hands meeting to brainstorm for keywords, then this could be a good next step.

You're not going to get everything you need from an all-hands meeting. This is just a starting point to get everyone thinking about keywords and questions. Ask people to email you more suggestions as they think of them, and schedule some breakout sessions for people or departments who are particularly motivated to help with keyword research.

You can also arrange to send out an email survey to everyone at the company. This is a fairly low-effort option, but you won't get as much participation as you would from a face-to-face meeting.

Beyond your current keyword research mission, this process also reveals the departments and people who care most about search traffic and website performance. You might consider forming a working group or committee for those who want to participate in future keyword reviews.

Marketers

Technically speaking, SEO is a form of marketing, so the people in the marketing department should have a great deal of insight for you, especially in terms of traditional marketing data. You should already be working with your company's marketers by this point unless you're an outside SEO consultant, in which case we advise you to engage with the entire marketing department – not just your point of contact. Someone might have done some of this research already and produced a keyword list that you can work with.

Ask if there are any upcoming product launches or marketing campaigns that you can align with your SEO efforts.

Be wary of letting marketingspeak and insider jargon slip into your keyword list. Marketers try to create a certain impression in customers' minds by using specific terms, but the language that customers actually use may be quite different.

Salespeople

No one talks to customers more than the people in sales. They know the exact language that customers use, the problems people are solving by buying a product, the most common questions people ask before buying, and related concerns such as price, reliability, warranty, returns, and support. Specifics are important in keyword research, which is why it's a good idea to talk to sales after you've talked to marketing. For instance, the people in marketing may have said that "affordable" is a good keyword, but the salespeople may say that customers most commonly say or respond to "cheapest" or "least expensive" instead. It's too early to make judgements, though, so for now include all three. You'll identify the highest-value keywords later, and filter out the duds.

It's also useful to ask the salespeople for feedback on the company's marketing efforts. Which parts of the marketing funnel are delivering qualified leads? Which aren't? Marketing is supposed to enable the sales team, but there can be some disconnection – even some resentment – between them.

Whereas marketing, user experience design, and product management may have formal personas or avatars that represent ideal customers, sales may have its own informal labels. For instance a frequent, big-spending customer might be referred to as a "whale," a past customer may be labeled as an "upgrader," or an existing one as a "renewal." Each of these classes may have entirely separate sets of keywords. If you inherited customer avatars from the marketing department, it would be a good idea to modify them to align with the sales perspective.

IT

The system administrator and/or webmaster will know if there is any internal data that might be useful to you.

In addition to the Web logs and analytics services that we've already mentioned, the IT department will probably also have access to records from your site's built-in search function. The on-site search queries of existing visitors and customers are a keyword gold mine. If your site doesn't have on-site search, then consider implementing the Google Programmable Search Engine: https://programmablesearchengine.google.com.

Support or Customer Service

The support department is more of a resource for content creation ideas than keyword research, but it's still useful to ask what words and phrases customers use when talking about the product. Like with sales, customers who contact customer service may fit into specific categories. If so, revise your avatars to account for this perspective.

Support personnel are a window into the problems and known issues with your products. As much as the marketing leadership would like to pretend

these don't exist, in reality your customers are probably searching for them, and your job is to direct those searches to good content, so you need to know what they are. As painful as it may be, you should (perhaps secretly) append the word *sucks* to your branded keywords. If people are searching for your product name and *sucks*, then you should be aware of it. You might also include other negative terms like *scam, unreliable,* and *ripoff,* or whatever might apply to what your business sells.

It's also a good idea to add *support* to your branded keywords. If your customers need product support, you want your support page to be found – and you certainly don't want your competitors' sales pages to rank above it.

Founders and Owners

They started this company for a reason, right? What's their story? What problem did they set out to solve? What's the next step for this company? What do they want this company's public image to be?

Customers

If you have access to existing customers, it's useful to ask them how they found your site. Some organizations include this question as part of the closing or checkout process. You can also set up a survey on your site, or as an email follow-up after the sale. Some people like to offer discounts for customers who complete surveys. While this will get you more survey responses, it'll also add a lot of junk and bias to the data, which could lead you astray.

Try to talk to representatives from each of your customer avatars.

Not everyone finds your site through a search engine, so customer feedback is not usually directly useful for discovering keywords. Indirectly, however, the process that a customer followed up to the sale can reveal valuable keywords that you wouldn't have expected. For instance, before buying a new video card from your PC retail site, a customer may have read a review on Tom's Hardware, then asked questions about it in the forum. The names of websites and publications peripheral to your business may make good

keywords, and discussion forums are an excellent source of relevant customer questions.

Non-Customers

Sometimes it helps to get a fresh, unbiased perspective. So far everyone you've talked to is an insider. What would someone outside of this industry search for if they wanted to buy from your site?

At some point, most or all of your customers were new to this concept or industry. What was their search journey like?

You can also get a broader perspective by looking up industry associations and media sites that pertain to your business. What language are they using when they talk about this subject?

External Resources for Keyword Research

By this point you should have a substantial list of raw keywords. The next step is to use third-party keyword research tools to find similar search terms. We covered the most important ones in Chapter 3, but there are a few niche keyword research tools that you should consider in addition to one or more full-service SEO platforms or data providers, especially if your company or client relies heavily on mobile search traffic and/or seasonal trends.

Researching Natural Language Questions

Natural language searches are important for all search methods, but they're critical for mobile search marketing. Compared to desktop search, people who search from a mobile device are far more likely to speak their query than type it, and when they do, they tend to use natural language questions or prepositions (statements that imply questions) instead of search-friendly keywords.

If you can, try to have at least one corresponding natural language question for every keyword in your list. Ideally you'd have a series of related

questions that refine the search scope; Google calls this process a *journey*, and it saves the context of each user's journey so that it can provide pre-refined results if they come back to it in the future. The idea here is that searchers aren't seeking a single, objective answer; they're subjectively evaluating many possible answers by following their curiosity and exploring a topic over a period of time. An efficient SEO strategy includes engaging with those searchers as early in their search journey as possible.

For instance, consider the previous example of keywords for smartphone accessories. Thinking about the domain and the topics within it, we came up with two good natural language questions that could lead to an accessory sale:

- What kind of rice to use to dry out a soaked iPhone?

- Is smartphone insurance worth it?

Let's work with the first one. Is this really the first question in someone's search journey? If we go back a step or two, some better starting points might be:

- Is the iPhone waterproof?

- What do I do with a wet iPhone?

- How do I dry out an iPhone?

- Will an iPhone work if you drop it in a pool?

The next level might be:

- What kind of rice to use to dry out a soaked iPhone?

- How much does it cost to repair iPhone water damage?

- Is water damage covered by AppleCare?

- Can I trade in a waterlogged iPhone?

Remember: the goal here is to sell a waterproof iPhone case. Some of these questions are reactive, and some are proactive. Either way, the people who ask these questions are probably interested in a waterproof iPhone case, even if they aren't directly searching for it yet. They may not even know it exists, in which case your content will shape their first impression.

Another great example is baby furniture and clothing. If that's what you're selling, then the search journey probably starts with:

- What are the most popular baby names?

- How much maternity leave do I need?

- How much does a full-term pregnancy cost?

At some point later in this search journey, the person who asked these questions will need a crib, crib accessories, a night light, a rocking chair, and baby clothes – things your site sells. They aren't searching for them right now; they will be, but they haven't got around to thinking about this topic yet, so this is your chance to get in front of this future customer ahead of your competitors.

Another path to consider is a parent who is having a second (or more) child. They might ask:

- Can I reuse my old crib for my new baby?

- Is my car seat still legal?

- Are used baby clothes safe?

If you're stumped, or if you want as many questions in your list as possible, consider the third-party tools in the subsections below.

> **NOTE**
>
> SEO platforms like RankRanger and Moz Pro are also excellent sources for natural language questions.

AlsoAsked.com

This site analyzes one of your natural language questions, and returns a mind map of all related questions in the search journey. You can drill down into each related question if the topic is too broad.

This service is free to use on a limited exploratory basis, but if you want to export to CSV (which you'll need to do in order to import the list into your spreadsheet), you'll have to upgrade to a paid account.

The AlsoAsked data comes from Google's "people also asked" SERP feature.

AnswerThePublic.com

This site analyzes one of your topics, and returns a mind map of the most popular natural language questions and prepositions within it.

This service is free to use on a limited exploratory basis, but the paid version includes historical reporting, "listening alerts" (notifications of changes for the topics you're tracking), CSV export (required for importing data to your spreadsheet), and extensive training materials.

AnswerThePublic gets its data from Google's autocomplete feature.

Researching Trends, Topics, and Seasonality

Hopefully your company's marketing department has already done plenty of research to define past trends and seasonality, and it's available to you to use. If not, or if you're launching an entirely new business, then this is a bit out of scope for an SEO project; seasonal market research and planning is a whole-company effort.

A trend is a pattern of increasing and/or decreasing activity that is supported by historical data. Seasonality is a trend that recurs based on well-defined and predictable conditions. If you have enough data, then both of these concepts are easily established with ordinary keyword research and analytical tools. (There's more detail on this topic in the "Trending and Seasonality" section later in this chapter).

Emerging trends are much more difficult to plan for because current events cause keywords to fluctuate on a daily basis. Google processes more than 5 billion queries per day, approximately 15% of which are new. It takes at least a day for search data to trickle into most keyword research sites, so there is a lot of variance and volatility with relatively new keywords, and regardless of how much traffic data you collect, at best you're getting an approximation of yesterday's search volume for any given keyword. Trends don't start or end on search engines, so if you really want to know what's trending today, you'll have to rely on other data sources. In addition to Google Trends (which was covered in Chapter 3), our suggestions are below.

Twitter

https://www.twitter.com

This is how Twitter describes itself: "Twitter is what's happening and what people are talking about right now." The sidebar tracks the current most popular trending topics, accounts, and hashtags, and the Explore feature has several sorting options for researching emerging trends.

Hashtags don't usually make good keywords, but they can inspire some ideas for good keywords and topics.

Soovle

https://www.soovle.com

When you type a topic or keyword into Soovle, it shows you the most popular related search queries on several sites, most notably YouTube, Amazon, Wikipedia, and Answers.com. People often search these sites because they want to learn more about something they heard about recently, and that's not necessarily what they'd go to Google to look for, so Soovle's result page may be more likely to show evidence of emerging trends before they show up in Google's keyword data (though Soovle does show Google and Bing results as well).

Even if you're not trying to capitalize on an emerging trend, Soovle is still an overall excellent source of inspiration for keywords and topics.

Keyword Valuation

Up to this point your keyword research has mostly focused on building your raw keywords and topics lists. The initial brainstorming phase is now complete. You've identified your topics, and collected a lot of potentially relevant keywords and questions, but you don't yet know which ones represent actionable and profitable SEO opportunities for your site. While they may be relevant, some may not be attainable (the cost of optimizing for them may be too high), and some may not have enough search volume (the benefit of optimizing for them may be too low).

The hierarchy of importance for keyword valuation is:

1. *Priority*: keywords that serve a major business objective, sales goal, branding initiative, or other critical marketing purpose should be considered above everything else.

2. *Relevance*: you only want to rank for keywords that are highly relevant to your site's content. Low-relevancy keywords are not necessarily bad, but they should be moved to a separate list in case you need them later.

3. *Popularity*: you only want to optimize for keywords that are used in a measurable number of actual searches. High-relevancy but low-popularity keywords should be filtered out (hidden in your spreadsheet), not deleted, because they may become popular in the future.

4. *Difficulty*: if the cost of acquiring traffic is higher than the benefit of converting it, then you're wasting money. High-difficulty keywords that are highly relevant to your site can be broken down into variations that represent more efficient opportunities.

To obtain these metrics, you'll combine search data from SEO platforms with your own subjective ratings for topics and keywords. If you have tens of thousands of keywords, then this is going to take a lot of time and patience – but that's part of the job, it's what you're being paid to do.

The subsections below offer more detail on each of these points.

Importing Keyword Data

Before you go any further, you must populate your Keyword Plan worksheet with useful search data. Good data isn't free – or if it is, then it has limited utility – so don't go cheap on this part. Bad data will lead to bad decisions that will waste time and money, possibly get your site penalized, and maybe get you fired. In SEO, it's better to have no data than bad data, and it's better to do nothing than to do the wrong thing.

Our suggestions on search data providers are in the "SEO Platforms" section of Chapter 3. You only need one good source, but if you have the budget for it, you may benefit from combining data from several different providers. Regardless of which service(s) you use, you must at least be able to upload your list of raw keywords to it, and export recent search data about those keywords to a CSV file with columns that include:

- Monthly search volume

- Keyword difficulty (also known as "keyword competition")

- CPC (optional, but nice to have, especially if a keyword difficulty score isn't provided)

- Current rank (if there is one)

- URL (of the page corresponding to the current rank)

If there are other metrics that you have good data for, add columns for them.

Once you have a CSV file with the exported search data, open your Keyword Plan spreadsheet, create a new worksheet tab and rename it to *Data Import* (or something that reflects where you got it from, such as "Semrush Import"), then import the CSV into it (using a spreadsheet function, or plain old copy-and-paste). Modify the headings and the order of the columns to match what's in your Keyword Plan worksheet, then copy-and-paste the data into it. Alternatively you may find it easier to re-order the headings in your Keyword Plan table to match the data export columns.

If you're a spreadsheet guru, you can write a formula that pulls data into each Keyword Plan column from the appropriate equivalent in the Data Import worksheet. Assuming column A is for keywords, and column B is for monthly search volume, then this formula would copy the search volume data from the Data Import worksheet to the same column in the Keyword Plan worksheet:

```
=VLOOKUP(A2, 'Data Import'!$A$1:$B$49995, 2,FALSE)
```

Repeat this process for all other data columns.

Evaluating Relevance

At a glance your entire keyword list might seem perfectly relevant to the topics you defined, but keep in mind: relevance is relative. A keyword might be highly relevant to your site's current content but have low relevance to your business model (or vice-versa). Seasonal keywords fluctuate in relevance throughout the year, and some branded keywords will become less relevant when products are retired or replaced with newer models.

Low-relevance keywords aren't necessarily lost opportunities; they're just not at peak value right now. When searchers click on your site and find the content to be valuable, they're likely to remember your brand, bookmark this page so they can return to it later, and potentially link to it when suggesting it to a friend. Low-relevance keywords, therefore, can present good opportunities to strengthen the branding of your site. This type of brand value can lead to return visits by those users when they are more likely to convert. Seasonal keywords also may be qualified as low-relevance outside of the season, but they could be highly relevant when laying the groundwork for the next season.

To identify currently relevant, high-quality keywords, ask yourself the following questions.

What is the search intent?

There are three kinds of query intentions:

Transactional :

> Someone is actively seeking to initiate a conversion (buy something, sign up for something, etc.).

Navigational :

> Someone is looking for a specific brand, product, or service.

Informational :

> Someone is seeking general information on a topic, or researching in preparation for making an upcoming (but not immediate) buying decision.

How closely is this keyword related to the content, services, products, or information currently on your site?

If you're selling smartphone cases, you may end up with some keywords that seemed like they might be relevant, but actually aren't. For instance, *iphone 11 data plans* doesn't apply to any of your product pages. The

search intent doesn't align with your current content. Perhaps you intended to write an article on this topic (which might be a great idea), but never got around to it. Under the right conditions this could be a relevant keyword, but presently it is not.

If searchers use that keyword in a query and click through to your site from a SERP, what is the likelihood that they will convert?

Conversion rate is directly related to how accurately your content matches searcher intent (though there are other factors as well). So if you do get around to writing that comparison review of various iPhone 11 data plans (with an extra section that explains how some iPhone 11 cases might affect the signal in positive and negative ways), you're reasonably well aligned with the searcher's intent. Traffic for that keyword may or may not convert – you'll have to keep a close eye on it.

How many people who search for this term will come to your site and leave dissatisfied?

If your title or snippet promises a comprehensive comparison review of iPhone 11 data plans, you have to deliver that or else the visitor will quickly leave. Don't take the position that more eyes on your site will necessarily lead to more conversions. You want to attract qualified leads. "Getting more people in through the door means more sales" is a brick-and-mortar retail store theory that relies on customers having to invest time and energy into physically travelling to your store, which doesn't translate well to the Web because it takes no effort to close a tab or click the "back" button.

Web analytics can tell you if you're delivering what you promise. A high bounce rate on a page says that your content did not meet people's expectations, though there could also be a usability or technical issue at fault.

Assigning Relevance Scores

Some keyword research tools offer their own proprietary "relevance" or "relevancy" score for search terms. This is not quite the same "relevance" that we're talking about in this section. Third-party relevancy scores are a measure of how many times a keyword appears in the overall search volume of its main parent topic or parent keyword (if you're researching similar terms). This can be useful for keyword research within a specific topic, and for finding related keywords, but it isn't a good metric to use for keyword valuation.

Instead of relying on a third-party score, think about your customer avatars and their search intents, your conversion goals, and the alignment between your keywords and your page content, then develop your own relevance scoring system. We suggest a scale from 1 to 3, with 1 being a highly relevant transactional query (likely to convert), 2 being a reasonably relevant navigational query (may convert), and 3 being a broad informational query (not likely to convert right now, but maybe later) or a seasonal keyword that is currently in the off-season.

Start at the top of your Keyword Plan worksheet and add a score to the *Relevance* column for all of your keywords.

Priority Ratings for Business Objectives

Similar to relevance scoring, you may also benefit from a subjective priority rating scheme to override or bias the standard cost-benefit ratio.

There are a lot of different ways to represent priority depending on the details of your business, and what you want to sort and filter by. Are there certain keywords you want to rank highly for no matter how difficult or costly? All other things being equal, are there certain topics or keywords that are a top priority for your business? If two keywords from two different topics have the same cost and benefit, and you can only focus on one, which would it be?

The best scoring method for this metric is one that closely matches your company's processes and culture. Many companies define an official hierarchy of *major business objectives* (MBOs) on a quarterly or yearly

basis. Marketing departments have a list of priorities as well, and sales departments often have monthly and quarterly targets for certain product groups or services. If nothing else, then you should at least have an idea of the relative benefit of each conversion type, or an estimate of how much each is worth. (If you only have one value for all conversions, then you probably don't need to assign priority ratings, and can skip to the next heading).

Using those sources, develop a basic rating system for your topics and keywords that reflect your company's current priorities. There are several ways to do this. We suggest one of these:

1. A simple Boolean flag (such as a different cell color, or the letter X) to indicate equally high-priority items; all unflagged items are equally low priority.

2. A numeric scale from 1 to 3 (or 1 to however many MBOs there are).

3. Approach it from a product management perspective – create a list of MBOs (using short names) and assign one or more of them to each keyword or topic.

In your Keyword Plan list, start at the top and assign scores or flags in the *Priority* column for each keyword (or just the topics, if that suits you better).

Filtering Out Low-Traffic Keywords

Low-traffic keywords are generally not worth the effort of optimization, but – like so many other things in SEO – it depends on the context and the specifics of your situation. There are reasonable scenarios where currently low-popularity keywords are worth targeting.

The first and most obvious one is to get ahead of an emerging trend, season, or future event. A lot of big events like the Olympics and the Super Bowl are planned many years in advance. Long before anyone knows who will be

competing in these events, you already know what they are, and when and where they will be held. Likewise, the best time to start optimizing for the next Back-to-School season is shortly after the previous one ends – when the competition for traffic has died down. (There's more on this subject in the "Trending and Seasonality" section later in this chapter).

You might also want to lay the groundwork for an upcoming product launch or marketing campaign. This can be tricky, though, if some of your keywords contain embargoed information.

As explained in detail earlier in this book, a long-tail keyword strategy involves targeting a high number of keywords that individually have relatively low traffic, but collectively offer a lot of opportunity. If that's the case, then try to figure out what the worthwhile traffic threshold should be. This is subjective, but there's definitely a point at which the numbers are so low that the majority of the measurable search traffic is probably just noise from automated queries that haven't been filtered out of the dataset. The irony of SEO data providers like Semrush and Ahrefs (and their upstream partners) is that their keyword data collection scripts are responsible for a certain amount of the search traffic that they're reporting. Even if they try to filter out their own queries, they can't always filter out their competitors'. Every time you test a query or examine a SERP, you're adding to those numbers as well. So in most cases, any keyword below 1,000 monthly searches should be a candidate for exclusion unless you have a good reason for keeping it. If you're unsure, and if the conversion potential seems high, then test it – the cost of ranking for a low-traffic keyword is intrinsically low.

Don't delete low-traffic keywords – filter them out instead, by using your spreadsheet's filter function on the monthly search volume column in your Keyword Plan table. If they're relevant, then they may become more valuable in the future.

Breaking Down High-Difficulty Keywords

Keyword difficulty can be calculated in a variety of ways, which is why each SEO platform has its own special scoring system for it. Typically it's a function of one or more of the following data points, weighted according to a top-secret proprietary formula: search volume, number of results, backlink count, and paid ad CPC. On their own, each of these is a weak indicator of keyword competition, but in collective they become more reliable, so you want to incorporate as many of them as possible when calculating your keyword difficulty score. You can do this on your own in your Keyword Plan spreadsheet, or you can rely solely on the difficulty or competition score from an SEO platform.

NOTE

Your data must be *normalized* (meaning all metrics are converted to the same scale) before you can use it to calculate a difficulty score. You can do this in Excel via the *STANDARDIZE*, *AVERAGE,* and *STDEV.P* functions.

High-difficulty keywords don't usually represent good opportunities. They're high-volume, but can be low-relevance; if they're too broad, they will bring in a lot of useless traffic with a high bounce rate and few conversions. If they are specific enough to generate qualified leads, then the cost of acquiring them is probably too high for the return on conversions.

Regardless, if you don't already have a salvageable SERP placement for a popular keyword (or sometimes even if you do), then it's going to be a costly battle to get into the top 10 (which is where 95% of SERP clicks are). It's much more efficient to break it down into more specific long-tail terms that are cheaper and easier to target, and hopefully represent a higher return in aggregate.

Going back to our shrimp example, let's say that *shrimp stir-fry* is one of the most popular search terms in your list, with several indicators of very high difficulty. You can reduce the difficulty by breaking it out into related keywords, phrases, and questions that are more closely associated with

specific products or content on your site (and are therefore more relevant as well):

- Shrimp stir-fry recipe

- Shrimp stir-fry recipe "low sodium"

- How to cook shrimp stir-fry

- How much fat is in shrimp stir-fry?

- Gluten-free shrimp stir-fry

- Best oil for shrimp stir-fry

- Wok for shrimp stir-fry

- What seasoning do I use for shrimp stir-fry?

> **NOTE**
>
> The best way to do this is through one of the SEO platforms we've recommended. You could do it on your own, but if you have access to good data, why aren't you using it?

Next, check for potentially valuable disambiguations and alternate spellings and phrasings. Misspellings, hyphenation, verb tenses, and singular / plural variations of high-difficulty keywords can be extremely valuable opportunities:

- Shrimp stirfry

- Stirfried shrimp

- Stir-fried shrimp

- Shrip stir-fry

- Shrimp stir-fy

A lot of English words are associated with unique colloquialisms or alternate spellings across different dialects, such as plow and plough, check and cheque, donut and doughnut, and hiccup and hiccough. In this case *prawn* can be another word for *shrimp*; you might try substituting *prawn* and *prawns* for *shrimp* in those keywords. Don't be afraid to consult a thesaurus!

When you've broken down a high-difficulty keyword into more efficient and affordable variations, be sure to evaluate them for relevance before adding them to your spreadsheet. Don't delete the original keyword – you'll want to keep track of that in the future, and it provides an excellent reference point for calculating opportunities for other keywords.

Trending and Seasonality

The keyword plan that you've been building has been focused on present-day opportunities with current data. If you happen to be doing keyword research on Black Friday or during a major industry convention where a lot of product releases and keynote speeches are generating buzz, then your data will probably be affected. That's why monthly keyword reviews are important (that's covered in more depth later in this chapter), and it's also why you should create separate keyword plans for seasonal search marketing. At the very least you should create a new column for seasonality, or a new worksheet that enables better sorting and filtering for multiple seasons without cluttering up your main Keyword Plan.

Many keyword research tools offer a fine-grained, long-term view of seasonal variance in topics and keywords, but the best overall big-picture tool for analyzing trends and seasonality is Google Trends. That's going to give you access to the largest amount of seasonal search data, and you can easily compare several related keywords to see if there are better opportunities in and out of season. Even a tiny difference like changing a singular word to the plural can yield different seasonal variance.

Literally any keyword can be seasonally affected, and not always during the times of year that you may expect. It's likely that there are regular and

predictable trends that you aren't aware of because they don't align with a traditional event or holiday, but could be valuable to your site. For instance sales of memory cards for digital cameras could spike slightly in March due to a combination of unrelated overlapping factors such as college basketball playoffs, Vancouver Fashion Week, and a yearly increase in cruise ship voyages. If your site sells memory cards, you might not even be aware of the impact of these trends because your competitors are optimized for them, or because you've always seen a steady increase in sales from March to August, and you assumed it was due to the usual peak wedding and vacation seasons (which are probably much more expensive to target). It's worthwhile to challenge your assumptions and investigate any potential trends that you find. Try to break them down to account for every major contributing factor.

Sometimes there are so many trends and seasons that impact your business that it makes sense to step back and focus on the big picture first. For instance a costume shop is busy year-round supplying local and travelling actors, theater production companies, makeup artists, models, musicians, clowns, and a wide variety of private parties and special events. Halloween might be a busier time of year than normal, but it might also be prohibitively expensive to optimize for the Halloween season due to broader competition and low conversion rate. If you specialize in expensive high-quality costumes and professional stage makeup, then it might not be worth the effort to compete against low-margin high-volume retail juggernauts like Wal-Mart and Amazon for sales of cheap Halloween costumes. School semesters might also be "seasons" because of school plays, ballets, and operas. The common theme here is that most of this costume shop's business is local, so even though the national Halloween season might seem like an obvious choice for seasonal optimization, it's likely that focusing on smaller local trends and non-Halloween customers is a more efficient and profitable effort.

Seasonal optimization is a year-round process, and thus so is seasonal keyword research. If you want to target the Halloween season (from August to October), you can't expect to start the process in July – or if you do,

don't expect to get very far with it this year. Search traffic tends to start going up two or three months in advance of major holidays and events, but Halloween costumes and Christmas gifts both draw heavily from pop culture fads, trends, movies, TV shows, and news events from the previous year, so there's no such thing as "too early" to start collecting keywords for the next holiday season.

Current Rank Data

In order to get where you want to go, first you have to know where you are. The SEO platforms we've recommended are capable of analyzing your site and providing the current SERP rankings for every page, as well as the exact URL that's being indexed at that rank. If multiple URLs are ranked for a given keyword, then use the highest-ranked one for now. (You'll decide what to do with the lower-ranked pages in future chapters.)

In order to calculate an opportunity score (which is covered in the next section), you'll need to have a number in every cell in the *Current rank* column. If there are highly-relevant high-priority keywords that you want to place for but currently don't, then fill in the blank with "101." A SERP rank of 101 is more or less equivalent to not ranking at all, and it will be above the "top 100 ranked pages" threshold that most data providers use as a standard report, so if you improve your rankings for these keywords in the future, you'll be able to see your progress.

It's possible to continue the keyword valuation process without current rank data (and if you're launching a new site, you'll have to), but that makes the cost (and therefore, opportunity) calculations less accurate.

Finding the Best Opportunities

At this point you have enough data to identify good opportunities for optimization. Use your spreadsheet's filtering, sorting, and pivot table functions to narrow the scope of your Keyword Plan so that it only shows the keywords that you want to work with. Particularly with pivot tables, it's best to create a new worksheet named *Opportunities* for this purpose.

That's all you need to do for now. If you want to go further, there are some extra valuation considerations in the subsections below.

Calculating an Opportunity Score

It's possible to create a new column to calculate an *opportunity score*. Unfortunately it's difficult to provide specific guidance on how to calculate this because we don't know exactly what data you have, what your goals are, or which factors are most important to you, so we don't know exactly what a "good opportunity" means to you. Throughout this chapter we've used mostly retail examples because they're common and easy to use for that purpose. However, we also have clients who aren't doing retail at all. Some are focused on branding, others are collecting qualified leads for sales calls, or acquiring email list subscribers. If you have a good idea of how much an average conversion is worth to you, though, then you can calculate a generic opportunity score by following the traditional cost-benefit paradigm:

Opportunity Score = (number of potential conversions * conversion value) / cost of optimization

If this is an initial or early-stage keyword plan, then you won't yet have a good method for calculating the cost of optimization. Some people use CPC as a cost metric here, and that's better than nothing, but it won't be very accurate. While organic search costs and paid search costs will scale similarly with keyword difficulty, you cannot reliably calculate one based on the other. However, you do know what your current rank for a keyword is (or you used 101 as a placeholder value for unranked pages) and how difficult it will be to compete for a higher rank, so here's how you might reframe "cost" and "benefit" in terms of effort and traffic:

Opportunity Score = ((relevance * priority) * search volume) / (difficulty * current rank)

This formula won't work as written because the data isn't normalized (the metrics are not on the same scale), and the various factors are not weighted according to importance. You can normalize the data with a spreadsheet formula (don't overwrite your Keyword Plan columns, though – use the Opportunities worksheet instead). Weighting of the metrics is entirely subjective. We suggest experimenting with different weighting to align with your budget and implementation timeline.

SERP Space Considerations

By default, Google has traditionally provided 10 organic search results per page (though users can modify their search settings to display longer SERPs), but that's not a rule or a guarantee. Organic search results can occupy a varying percentage of SERP real estate due to encroachment by Google Ads and *SERP special features*:

Knowledge Panel :

> This is a variable-sized area on the right side of the SERP that pulls relevant content from the Google Knowledge Graph. Often an informational query is satisfied by a Knowledge Panel, which means there's no clickthrough to a result.

OneBox results :

These are trusted answers to search queries. They're shown in a box above the organic results, and are usually short text excerpts. Some examples (among many) of queries that generate OneBox results are: word definitions, unit conversions, package tracking, mathematic equations, health issues, and hotel searches.

Featured snippets :

Similar to a OneBox result, except the information comes from a highly-ranked and trusted site instead of from the Knowledge Graph. This is a much larger text excerpt than a regular snippet, and it appears above the URL instead of below it. You're more likely to see featured snippets for natural language queries and mobile searches.

Map pack :

This is a small excerpt of an image from Google Maps that can appear in the middle of the organic results, along with three top Google My Business listings. This will appear for local (or location-specific) queries such as *sushi restaurants in Orlando*.

Sitelinks searchbox :

This is a search field that sometimes appears below a snippet. If Google determines that a broad query is likely to lead to a second query on a site that has its own integrated search engine, then Google will try to pre-empt that second search by providing a search field with a site-limited scope. This often happens when someone uses a URL or a well-known brand or site name as a search query. For instance, if someone searches Google for *pinterest*, Google will provide a sitelinks searchbox under the top result for pinterest.com.

Rich results :

In lieu of a snippet, some search results can display an image thumbnail or review star rating. This is typically for results where the rich element

is an important part of the content, such as pages that contain reviews of books or movies, or interviews with celebrities.

Carousel :

If there are multiple pages on a site that are similar and contain rich results, Google may choose to display them in a horizontal carousel.

Enriched results :

When search results for job postings, recipes, or event listings lead to pages that contain interactive elements, Google may add some of that functionality to a rich result.

The more Ads and SERP features there are for a keyword, the shorter the list of organic results on the first page, and the less opportunity you have for a clickthrough to your site. You don't want to spend time and money improving your rank on a SERP that has a low organic CTR. Special features and ads can reduce the number of page 1 organic results, and siphon off much of the traffic from the remaining page 1 results.

Special features also occasionally create bizarre scenarios in which you will get less traffic by ranking higher. If Google is excerpting your content with a special feature, then searchers probably won't need to click through to your page. By de-optimizing that page so that it ranks slightly lower, you can force Google to excerpt someone else's content, which allows your page to return to the normal SERP list (and hopefully stay on page 1).

This encroachment can apply to any keyword in your list, but it's especially impactful on natural language questions and local keywords. Keep in mind, though: it's only a major issue for pages in the top 10 results. You can easily check SERPs for special features by querying Google with your keywords, but most of the good keyword research tools have scoring systems that measure organic CTR and/or SERP feature encroachment.

Rank Threshold Values

When calculating the costs and benefits of optimization, it helps to adopt a broad hierarchical view of search rankings. The actual rank numbers don't matter very much when they're within certain ranges. Ranks lower (meaning a larger number) than 20 are only meaningful from the perspective of monitoring the impact of your SEO efforts.

Above all else, rankings aren't worth spending time on until the fundamentals are solid. If your site recently launched, or if it has technical or UI problems that prevent a lot of your pages from being indexed, then you don't have any data to work with yet, and the first goal is to get everything into the index at any rank. Likewise if you're repairing a site that is being penalized for spam or black-hat SEO tactics, then the first goal is to clean it up and get back into the index. Keyword research is still important, but don't worry about rankings yet.

From unranked, the next threshold is the top 100. From a conversion standpoint the low end of this threshold is meaningless because the 9th page of search results is the SERP equivalent of Siberia; it may be on the map, but no one goes there except by accident or adventure. The 10th page of results is the doorway to the Twilight Zone; results beyond that are not truly ranked, and Google will only provide an estimate as to how many there are unless you navigate to the last result page (whatever it may be). Regardless of traffic or conversion rate, from an analytics standpoint it's worth the effort to get into the top 100 because most SEO platforms only provide data for the top 100 sites for each keyword, so you'll have access to useful metrics and can start tracking your progress.

To get minimal value out of a keyword, you have to place within the top 20 results. Only 5% of organic clicks go to the second page (usually results 11 through 20, but not always), so if that's where you are then there's a faint hope of a clickthrough, but unless the search volume is very high, it may not be measurable or predictable. So to continue the analogy: if you're #21 you're still in Siberia, but if you've steadily improved your rankings to get there, then you can start to get an idea of what kind of effort it will take to break into the top 10. More importantly, your SEO tools will be able to give

you better cost predictions, backlink suggestions, content ideas, and a more accurate competitive analysis.

95% of organic SERP clicks are on the first page, and the majority of those go to the top result. The further you are from #1, the fewer clicks you get, and remember: SERP features can kick you off the first page even if you're in the top 10, so if you want a high-priority keyword to be on the first page, you need to know what the threshold is for it. It would be disappointing to spend a lot of resources to get to #10, only to find out that a rich result kicks you back to page 2.

Take these thresholds into account when you're considering keyword valuation and planning. If you're currently placing #90 for a high-difficulty, high-volume, high-priority, high-relevance keyword, it probably isn't much of a cost to improve that until you start to approach #20.

Filtering by Topic

Instead of looking for the best opportunities among the entire keyword list, you might consider filtering them by topic instead, especially for a long-tail strategy where opportunity is measured in aggregate.

If the list is only a few thousand rows or less, you could sort your Keyword Plan list by each individual topic to get a quick impression of which topics represent the best opportunities, but it would be more efficient (and becomes a necessity when working with larger lists) to create a new worksheet that aggregates the number of keywords in each topic or superset, and their overall search volume.

Keyword Reviews

After you've developed a good keyword plan, you should schedule regular reviews to update the data and adjust your calculations. We suggest monthly reviews, since that's the interval that most search data providers use for updating keyword information.

You should also do a keyword review if there are any significant content additions to your site, changes in business policies or practices, or a shift in trends or standards. If the marketing plan changes, or if the CEO declares that they're going to "bet the company" on this new product release, or when old products are retired, or when a competitor announces a new product, do a keyword review.

Even a news story that barely relates to your products or content can impact the search volume of your keywords, such as the announcement of a new version of Microsoft Windows, political unrest in Taiwan, or a shipping container shortage. Anything that affects what people search for will affect keywords.

Sometimes a change can sneak up on you. Computer monitors, for instance, have switched connection standards a few times over the years, from VGA to DVI, then to HDMI, and more recently to DisplayPort. During each transition, people steadily search for converters that enable the outgoing standard to work with the incoming one. Let's say your site sells cables for electronics, and you have a page optimized for *hdmi to displayport adapter* that has been performing well ever since DisplayPort was introduced to the market. Your previous keyword plans have repeatedly identified this as a measurably better opportunity than *displayport to hdmi adapter*, but there will be a certain point in the transition between these standards when a fresh keyword review will show that *displayport to hdmi adapter* offers a much better opportunity because all new devices now have a DisplayPort connector, but everyone still has a bunch of old HDMI cables that they'd like to reuse. The sooner you know about that inflection point, the quicker you can respond to it with new optimization.

Resist the urge to say: "But I tested that, and I found the best query to optimize this page for." Optimization is a process, not an event. A simple change in the word order of a query can have a significant impact on traffic at any time, even if the context is the same. Though there may be a lot of overlap in the results between *hdmi to displayport adapter* and *displayport to hdmi adapter*, and even if they refer to the exact same product, they still lead to two distinct SERPs. Even if your page is ranked at the same position

in both SERPs, special features such as "People Also Asked" will favor the result that most closely matches the exact word order of the query regardless of its rank (within reason).

SEO platforms like the ones we've recommended will usually help you keep track of your progress over time, but you might find some value in analyzing the numbers on your own. We suggest creating a new keyword plan spreadsheet (by copying the old one, or starting from scratch) every time you do a keyword review. That way you can go back to previous plans and get a page-specific view of how a particular keyword's performance has changed.

Conclusion

Keyword research is a complex and time-consuming aspect of search engine optimization, but the rewards are high; once you learn where the keyword search volume is, you can begin to think about how that affects the information architecture and navigational structure of your site – two critical elements that we will explore in greater detail in Chapter 6. You'll also understand where your content gaps are, and how to alter your metadata to improve search traffic and increase conversions.

CHAPTER SIX

SEO Analytics and Measurement

An essential component of digital marketing is the tracking, measurement, and analysis of various data points via analytics and other data-oriented tools. These tools include traffic and performance analysis platforms, link analysis and discovery tools, keyword and topic research tools, and other systems designed to support your online marketing efforts.

Once you have selected the appropriate analytics and measurement tools to use in coordination with your SEO efforts, it is important to establish baseline measurements - with proper baselines accounting for as many variables as are applicable to your business and industry. These baselines can include metrics such as conversion rates by acquisition channel,

conversion rates by content type, URLs sending organic traffic, indexing status, backlinks, and more:

- Quantifying organic search traffic by search engine

- Quantifying a breakout of what site content areas are getting the current organic search traffic by search engine

- Quantifying conversion, event, and other trackable metrics by search engine

- Identifying poorly performing pages (e.g., low converting, high bounce rate)

- Identifying best-performing pages (e.g., high volume of traffic, high conversion rate)

- Tracking search engine crawler activity on the site

- Determining the number of indexed pages

- Determining whether the indexed pages are getting search traffic

- Determining whether best-selling product pages are indexed and getting search traffic

Your baselines are going to help you as you measure your efforts to generate various outcomes through SEO, including:

- Transactions

- Subscriptions and Memberships

- Lead generation

- Ad revenue generation

- Audience building / community creation

- Affiliate revenue generation

Why Measurement Is Essential In SEO

It is important to recognize that effective SEO requires understanding various SEO-related metrics, such as organic traffic, average position (as can be seen in your website's Google Search Console data), crawling, indexation, and backlinks – as well as understanding the relevant business outcome metrics (e.g., conversions, sales, and ROI). As an SEO practitioner, you should always aim to integrate understanding and analysis of each of these aspects in order to appropriately align your SEO efforts with your business's digital marketing goals, and to effectively measure the impact of your efforts.

Analytics Data Utilization: Baseline, Track, Measure, and Refine

An effective measurement process generally includes the following components, some of which may vary and/or be specific to your unique situation:

1. Determine your baselines based on your overall strategy for achieving your objectives. If your strategy for generating more e-commerce sales is to create more rich snippet results in Google, establish baselines appropriate to that strategy. Ensure you don't get false baselines due to seasonal factors or some other unusual events. Comparing year-over-year data will usually eliminate fluctuation due to seasonality. However, you must also consider how changes in the market, new competition, elimination of competition, industry consolidation, changes in your business strategy, and changes in the search engines themselves may affect your data.

2. Identify the appropriate tools for tracking and measuring the various metrics specific to your objectives and collect the newest data for each metric you decided to focus on. SEO can take many months (and usually longer) to show results - ensure you wait long

enough for your efforts to have a measurable impact. Some considerations:

- If your site is brand new, it may take longer for your changes to take effect.

- If the scope of the change is drastic (such as a complete redesign incorporating new information architecture, new URLs, etc. vs. a simple visual re-skinning), the time to see results will probably be longer.

- Sites that get crawled at great depth and frequency will probably yield visible results faster.

3. Compare the baseline data to the new data within your various data sources to identify areas for improvement, modification, and expansion, and refine your efforts.

If you don't see any change in your traffic, conversion rates, or other relevant metrics over the appropriate measurement timeframe (or if the changes are measurably negative), then you will need to examine the data more closely. Perhaps you've overlooked something important, or there was an error in your implementation. If you are achieving great results – such as a noticeable increase in traffic and/or conversions – then consider expanding your efforts and applying these effective strategies to more areas of your site.

Measurement Challenges

It might sound easy to record a set of SEO metrics before the start of the project and then compare the same set of metrics over 30, 60, or 90 days to measure the progress. But what if you don't make any changes during one quarter, and the metrics reflect an improvement that can reasonably be attributed to earlier SEO efforts? Conversely, what if the improvements are because of external business factors? How can your SEO project get attribution for its business impact?

One other issue which may significantly impact your ability to establish a baseline for measurement is a seemingly spontaneous drop in traffic. When analyzing the organic traffic to a site, if you notice a large drop in organic traffic it's vital to determine the cause of this drop before establishing your baselines. Large traffic drops can be caused by a number of factors including a large-scale site redesign or rebuild, a shift in the nature of the business as noted above, seasonal factors (which can usually be determined by looking at several years of data at a time), or possibly by organic search algorithm updates.

An extremely useful tool to use in determining if a large traffic shift might be caused by an algorithm update is the Panguin (a concatenation of the words panda and penguin) Tool created by Barracuda Digital (https://barracuda.digital/panguin-tool/). This tool allows you to overlay your Google Analytics organic traffic with Moz's history of algorithm updates to see if traffic shifts coincide with the updates. It's not always possible to definitively determine if you have been impacted by an algorithm update as occasionally several updates happen over a very short period of time or you may have a traffic drop due to another cause, which simply happens to coincide roughly with the algorithm update. However, before embarking on an SEO project it's important to note the possibility of how an algorithm update may impact the potential success of the project.

Analytics Tools for Measuring Search Traffic

The list below contains commonly used web, data, and customer insight analytics platforms for digital marketers:

- Google Analytics 4* (Google Marketing Platform: https://marketingplatform.google.com/about/analytics/) – *Note that as of July 1, 2023, Google replaced Universal Analytics with Google Analytics 4 (GA4), meaning all websites using the old GA tracking code no longer have their data being tracked as of July 1, 2023.

- Adobe Web Analytics (https://business.adobe.com/products/analytics/web-analytics.xhtml)

- Adobe Marketing Analytics (https://business.adobe.com/products/analytics/marketing-analytics.xhtml)

- Woopra Customer Journey Analytics (https://www.woopra.com)

- Clicky Web Analytics (Privacy-Friendly) (https://www.clicky.com)

- Matomo Web Analytics (Privacy-Friendly) (https://matomo.org/)

- Webtrends Analytics for Web Apps (https://www.webtrends.com/products/analytics-for-web-apps/saas/)

Valuable SEO Data in Web Analytics

You can extract numerous valuable data from web analytics. Below are some examples of information you may want to extract:

Organic traffic by search engine

A high-level data point is to identify your traffic by search engine. Generally speaking, if you are invested in SEO, you will likely see more relative traffic from Google than from other search engines because of Google's dominant search market share in most countries (though there will always be outliers or you may live in one of the few countries where Google doesn't dominate the market).

Organic impressions, traffic, and conversions by landing page and content type

An indirect way of measuring a component of your SEO efforts (including indexation status) is to keep an eye on the number of pages and types of content that are earning organic search impressions and traffic (generally measured as sessions, in most analytics applications). This data represents a

subset of the total amount of content indexed, but has greater value in that it indicates your content was compelling and ranked highly enough that users decided to engage with it.

There are several ways in which someone knowledgeable in analytics can configure your analytics account(s) to provide insight into which of your pages and content types are receiving organic traffic where the keyword is "not provided" (which will be the great majority of them). From this you can extrapolate which groups of related topics those landing pages are optimized for and which queries they actually rank for. You can also use Google Search Console to cross reference your pages and see what search terms are the source of that "not provided" traffic.

Referring Domains, Pages, and Sites

Referrer data is very useful to help SEO practitioners identify where non-organic traffic is coming from, which can often point to external links gained. You can often see those new links in these reports first, even before your backlink analysis tools report them.

Event Tracking

Event tracking goes one step deeper than basic analytics, and the ability to identify specific user actions (e.g., events), beyond transaction data, is incredibly valuable to digital marketers. Rather than simply observing what pages are visited and how many unique sessions are generated, event tracking allows you to narrow down groups of visitors based on the behaviors they perform on your site and how they engage with your content. Below is a basic list of trackable user behavior events:

Add to cart

> Studies have shown us that users who "add to cart," even if they do not complete the checkout process, are more likely to return to make a purchase. This is also a good way to calculate shopping cart abandonment and make changes to refine and improve the process.

Complete checkout

An obvious one; this action will show you what percentage of each user group is converting into sales. It is also of interest to measure what percentage of people start the checkout process but do not complete it.

Save to wish list

E-commerce sites offering wish lists are still in the minority, but wish lists are a great way to track interest that isn't quite a purchase.

Shares/send to a friend

Many sites offer a "share this page" function, and it is a great action to be aware of. If folks are sending out your link, you know you have a hit.

Subscribe to newsletter

A subscription is a tacit endorsement of your brand and a desire to stay in contact. It may not be a conversion, but for B2B, it may be the next best thing.

Contact form submission

Filling out a contact form can be even more valuable than newsletter subscription, in some cases. Though some of these forms will report support issues, many may contain questions about your products/services and will indicate a desire to open a sales conversation.

Email link

As with contact forms, direct email links have the possibility of becoming a sales contact. The best thing you can do is clearly label sales emails and track them separately from support or business issues.

Post comment

Anyone who is contributing content to the blog or participating should be paid attention to (as should those channels that earn user

engagement).

Social bookmark/share

All those folks who are submitting your content to Facebook, Twitter, Delicious, StumbleUpon, Reddit, and other social media and news aggregation/discussion sites deserve to be recognized (and sought after).

Register as a user

Registered users provide you with some form of information you can use in your digital marketing efforts, ranging from name and location to email address and phone number. This data can be useful in all of your digital marketing efforts, including email marketing and display retargeting.

Sign up for a subscription and/or membership

Many sites operate on the subscription or membership model, whereby recurring payments drive value beyond the original event.

Contribute content

When a user publishes content to your site, discovering the path the user takes in the process is important and can help you refine your site's layout and calls to action.

Add comment

As in the preceding list item, comments are a great predictor of engagement.

Vote/rate

Even low levels of participation, such as a rating or a vote, are worth tracking when every piece of participation counts.

Social engagement metrics

Likes, shares, mentions, follows, and many other user engagement actions are trackable.

Connecting SEO and Conversions

As we discussed previously in this chapter, it is important to tie your SEO efforts to the results they bring to the business. A fundamental piece of that is measuring the conversions driven by organic SEO traffic.

Below are some of the most common types of conversions (you'll notice that these are also trackable "events" as described above):

Sales and transaction

This is the one that everyone assumes is part of conversions. Sales and sales revenue (or better still, margin) conversions can be the simplest things to track, except when you are selling many different products at different price points and in different quantities. In this case, the process would need to be a bit more sophisticated.

If your site is advertising-driven, you need to look at the impact of organic search traffic on advertising revenue. If you have no financial goals for your site, you need to look at some of the other types of conversions and determine their value or worth.

Newsletter signups

Anytime a user signs up to receive regular communications from you it is a win. Even though there are not direct financial consequences to this it is still a conversion. Someone who has subscribed to something you offer is more likely to become a customer than a first-time visitor to your site, so you need to credit this type of conversion.

Subscriptions and memberships

Perhaps you offer content on a subscription or membership basis – each new subscriber or member is a converting user.

Content downloads

Many sites offer free downloads, such as white papers, or free downloadable tools. Even if you do not require a sign-up of any type, you should still count this as a conversion. You are getting your message out there with the downloads you offer.

Contact form submissions and phone calls

This is when your users request contact from you or contact you directly. Phone calls can be tracked through various tracking systems in available advertising platforms.

Users who share your content

This conversion happens when a visitor shares the information she found on your site with someone else. This includes Facebook Shares, and Links on Twitter. In addition, if your site has a "share" or a "send to friend" feature, you can track conversions by noting each time a user used that feature.

Users who link to your content

A user who links will visit your site and find its content useful, entertaining, or otherwise compelling enough to link to it from her own site.

It is important to assign a value to every type of conversion you receive. SEO software packages such as Conductor's Organic Marketing Platform (https://www.conductor.com/platform/overview/), Brightedge (https://www.brightedge.com/products), Searchmetrics (https://www.searchmetrics.com/suite/), and seoClarity (https://www.seoclarity.net/platform/) allow you to view search ranking data together with traffic and revenue data. This enables you to connect organic search traffic to conversions to measure SEO performance. You can

see sample screens from Conductor's Organic Performance Data for Content in Figure 6-1:
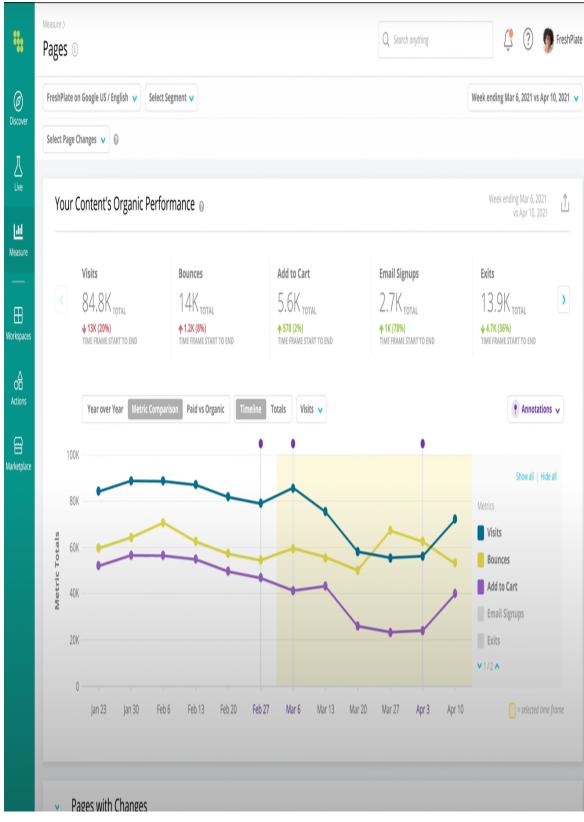
Figure 6-1. Conductor's Organic Performance data for Content

Attribution

Another issue to be aware of is attribution - identifying user actions that lead to specific outcomes. For example:

- A user performs a search, clicks on an organic search result, views a few pages of content on the web site, and leaves. The next day, she remembers what she read, performs another search, and then clicks on a paid search ad for the site and buys a product. In this instance, organic search should receive some credit (e.g., attribution).

- A user performs a search on a mobile browser, clicks on an organic search result, and leaves. A few days later, she remembers the website from the search results and downloads the mobile app for the company from that search result, and makes a purchase. Organic search should also receive some credit for this transaction.

Ideally, you should attempt to track the impact of your SEO efforts in a multi-channel data tracking environment. The following marketing attribution tools offer various ways to measure attribution across various channels:

- Adobe Analytics Attribution IQ (https://business.adobe.com/products/analytics/attribution.xhtml)

- Branch (https://branch.io/attribution/)

- C3 Metrics (https://c3metrics.com/)

- Attribution (https://www.attributionapp.com/)

- Dreamdata (https://dreamdata.io/)

Segmenting Campaigns and SEO Efforts by Conversion Rate

Once you have conversion and attribution tracking configured, how do you use it to focus your SEO efforts? One of the things you will want to do is to track conversion data in different ways:

Conversions by keyword

> What keywords are bringing the best results?

Conversions by referrer

> Which traffic source is offering the best conversion?

Conversions by web page

> Which pages on your site result in the highest number of conversions?

Conversions by initial entry page

> Which initial entry pages ultimately lead to the highest number of conversions?

Increasing Conversions

As an SEO practitioner, you should strive to become an expert at conversion optimization, because higher conversion rates mean higher impact for your SEO efforts. There are various tools available to assist with conversion rate optimization (CRO), including:

- Optimizely (https://www.optimizely.com/)
- VWO (https://vwo.com)
- Mouseflow (https://mouseflow.com/)
- Plerdy (https://www.plerdy.com)
- Attention Insight (https://attentioninsight.com)
- User Testing (https://www.usertesting.com/)
- HotJar (https://www.hotjar.com/)

- BuzzSumo (https://buzzsumo.com)

Calculating SEO Return on Investment

An effective SEO process is one that continuously works toward a positive return on investment. Your methodology for determining the ROI of your SEO will be very specific to your business, though most will generally include the following components:

Number of people searching for your keywords

> This can be challenging to estimate, because you cannot completely map out the long tail. One rough estimation strategy is to multiply the search volume for the top terms for your business by 3.3 (i.e., assume that the head terms are about 30% of the available volume).

Average conversion rate

> Once you have attracted the user through organic, how successful are you at completing a conversion? Typical conversion rates for a website might be between 2% and 5%. It should be easy to get this data from your analytics. You should already know what your conversion rate is!

Average transaction value

> Last but not least, factor in the average transaction value. Again, this is data you already have.

An example formula for calculating ROI:

SEO Revenue = People searching for topics related to your content (e.g., identified keywords and/or queries) * click-through rate * average conversion rate * average transaction amount

For example: 10,000 per day * 10% * 5% * $100 = $3,000 per day

SEO ROI = SEO Revenue / SEO Cost (use total $ spent for salaries, consulting, and development, and/or number of hours spent)

For example: $3,000 per day / $500 per day = an ROI of 6X.

An alternate theory and calculation

Predicting an SEO project's ROI based on impressions and click-through rate can, in some cases, be problematic to address because you have very little control over the variables. You end up relying on numbers that you have a limited ability to influence. As an alternative approach, you can measure and track SEO ROI based on an increase in search visibility.

To do this, begin by determining two things:

- How many pages and content types are generating traffic for a given time period

- How many visits does each page and/or content type receive for a given time period

Next, record these supporting metrics:

- Average ranking across the whole keyword spectrum

- Average click-through rate

Now, by making it easier for search bots to find more pages, consolidating duplicate content, improving the depth and breadth of content, changing your pages to better match your user's needs, and improving page titles, metadata, and microdata, you should see an increase in the number of pages getting search clicks and/or the number of clicks per page. The combination should result in more traffic when compared year-over-year.

Modifying the formula above to account for this alternate approach:

SEO Revenue = Increase in (pages/content getting search clicks * search clicks per page) * average conversion rate * average transaction value

Diagnostic Search Metrics

Thus far in this chapter we have focused on the basics—the dollars and cents of determining whether you have succeeded in your SEO campaign. As we noted at the beginning of the chapter, these should be your first priority in your analytics efforts. In this section, we will start looking at metrics that you can use to diagnose specific SEO issues. An example of this would be finding out whether a major section of your site is not indexed.

Another example is seeing how your traffic growth compares to that of your competitors (this helps you decide whether you have set the right objectives for your efforts). Numerous tools allow you to monitor your site and those of your competitors, providing insight into your SEO progress. You can also use these tools to figure out what your competitors are doing from an SEO perspective. This type of intelligence can provide you with new ideas on how to adapt your strategy to get better results.

As with all such tools, it is always important to understand the context of the tools and to have an idea as to what you are looking for. Better knowledge of your competitors' strategy is certainly one valuable goal. Detecting a problem in how your website is crawled is another.

By selecting specific and actionable goals, you can set yourself up for the highest possible return.

Site Indexing Data

It is valuable to know how many pages in your site are in a search engine's index. This is important for two reasons:

- To determine whether important parts of your site are not in the index. If key parts of the site are not in the index, you can embark on an effort to determine why.

- Looking at competitors' indexed pages can help you learn about your competitors' sites and strategies.

You can get an estimate of the number of indexed pages for a site using the site:yourdomain.com command in Google or Bing. Figure 6-2 uses the results for the New York Times web site.

Google   site:www.nytimes.com   X   Q

Q All   🖼 Images   📰 News   🛍 Shopping   📍 Maps   ⋮ More   Tools

About 4,890,000 results (0.18 seconds)

https://www.nytimes.com   ⋮

The New York Times - Breaking News, US News, World News ...

Live news, investigations, opinion, photos and video by the journalists of The New York Times from more than 150 countries around the world.

https://www.nytimes.com › ...   ⋮

The New York Times Canada - Breaking News, US News ...

The New York Times provides live news, investigations, opinion and video from the United States, Canada and around the world. Our 1700 journalists report on ...

https://www.nytimes.com › wirecutter   ⋮

Wirecutter: New Product Reviews, Deals, and Buying Advice

Wirecutter tests and reviews the best tech, appliances, gear, and more. You can trust our veteran journalists, scientists, and experts to find the best ...

https://www.nytimes.com › saved   ⋮

Reading List - The New York Times

Save New York Times content to read later, when you have time, on your computer, Mobile phone or Tablet.

https://www.nytimes.com › tufts   ⋮

Tufts University - Access NYT

Log in with your institutional credentials. Create a NYTimes.com account using your institutional email – or use your pre-existing profile. Get your access now:.

http://www.nytimes.com › FIU   ⋮

Access NYT « The New York Times in Education

Florida International University - Miami, FL Instructions to Redeem Complimentary NYTimes.com Subscription. NYTimes.com Access Step by Step Directions.

Figure 6-2. Site indexing data

As shown in Figure 6-3, Google reports over 4.8 million pages in the index. While this is not necessarily an exact figure (more accurate Google indexation data for your site is found in your site's Google Search Console data, a free tool described in the next section), many site owners can use this information to estimate the number of their site's pages that are indexed. This tool can also be used to identify the indexation status of specific pages, such as those containing a particular term in the URL:

Google

site:searchengineland.com inurl:privacy   ✕   🔍

🔍 All   📰 News   📖 Books   🖼 Images   ▶ Videos   ⋮ More   Tools

About 165 results (0.30 seconds)

https://searchengineland.com › Channel › PPC   ⋮

Why consumer privacy is Google's ace in the hole - Search ...

Nov 23, 2021 — Google's ad business changes mean that advertisers should prepare and tap
into their first-party data to stay ahead.

https://searchengineland.com › library › privacy   ⋮

Privacy - Search Engine Land

Privacy news from Search Engine Land provides information, trends and how-to's marketers
need to stay ahead of the competition.

https://searchengineland.com › Channel › PPC   ⋮

Google's Privacy Sandbox ad technology testing begins

Mar 31, 2022 — Developers will gain access to Privacy Sandbox's newest measurement
proposals: Topics, FLEDGE and Attribution Reporting.

https://searchengineland.com › Channel › PPC   ⋮

Google's Privacy Sandbox, including the Topics API, is coming ...

Feb 16, 2022 — Google has unveiled plans to roll out several Privacy Sandbox initiatives on
Android, including the Topics API, the FLEDGE API for custom ...

https://searchengineland.com › ... › Amazon Advertising   ⋮

The state of tracking and data privacy in 2020 - Search Engine ...

Feb 14, 2020 — Here's where search marketers find themselves in the current entanglement of
data and privacy and where we can expect it to go from here.

https://searchengineland.com › Channel › SEO   ⋮

Edward Snowden calls out Google over search engine's privacy

Nov 17, 2021 — Will this tweet trigger some change at Google? Probably not - but it might be
wishful thinking.

Figure 6-3 shows that the http://www.searchengineland.com domain has 165 pages in the index that have "privacy" in the URL.

Index-to-Crawl Ratio

This is the ratio of pages indexed to unique crawled pages. If a page gets crawled by Googlebot, that doesn't guarantee it will show up in Google's index. A low ratio can mean your site doesn't carry much weight in Google's eyes.

Search Visitors per Crawled Page

Calculated for each search engine separately, this is how much traffic the engine delivers for every page it crawls. Each search engine has a different audience size. This metric helps you fairly compare the referral traffic you get from each.

As you optimize your site through multiple iterations, watch the aforementioned KPI's to ensure that you're heading in the right direction. Those who are not privy to these metrics will have a much harder time capturing the long tail of SEO.

Free SEO-Specific Analytics Tools from Google and Bing

To get data that is specific to your website directly from these two search engines, you can use their free tools provided for website owners and webmasters, Bing Webmaster Tools (BWT) (https://www.bing.com/webmasters/about) and from Google Search Console (GSC) (https://search.google.com/search-console/about).

Both of these tools are very useful, and the data they provide can assist you tremendously in understanding how your website is performing in these search engines, and why.

Using Google Analytics (GA) and Google Search Console (GSC) Together

Google provides for data sharing between GA and GSC to enable you to correlate GA data with organic search data in a more streamlined fashion, by enabling correlation between pre-click data (e.g., search queries, impressions, and average position) and post-click user behaviors (e.g., pageviews, bounce rate, and conversions).

Differences in How GA and GSC Handle Data

There are some differences between how Google Analytics and Google Search Console handle and present data. For example, GSC aggregates Landing Page data under canonical URLs (defined below), whereas Google Analytics uses the actual URL.

Canonical URL: A *canonical URL* is the URL of the page that Google thinks is most representative from a set of duplicate pages on your site. For example, if you have URLs for the same page (`example.com?dress=1234` and `example.com/dresses/1234`), Google chooses one as canonical. The pages don't need to be absolutely identical; minor changes in sorting or filtering of list pages don't make the page unique (for example, sorting by price or filtering by item color).

The canonical URL can be in a different domain than a duplicate URL.

Figure 6-4. (Source: https://developers.google.com/search/docs/advanced/crawling/consolidate-duplicate-urls)

As a result, GSC will show aggregated impressions and aggregated clicks in their reports, as described below:

URL	Impressions	Clicks
http://www.example.com	1000	100
http://m.example.com	1000	100
http://www.example.com/amp	1000	100
Canonical URL	Aggregated Impressions	Aggregated Clicks
http://www.example.com	3000	300

Most of the time, landing page and canonical URL are the same. You can use the Search Console Query Inspection Tool to identify the Canonical URL associated with a landing page.

Figure 6-5. (Source: https://support.google.com/analytics/answer/1308626)

Differences Between Metrics and Dimensions in GA and GSC

The following table identifies terms that are used in both Search Console and Analytics reports.

Term	Search Console usage	Analytics usage
Impressions	Used for both Google Ads impressions and Google Search impressions	Used exclusively for Google Search impressions
Clicks	Used exclusively for Google Search clicks	Used for both Google Ads clicks and Google Search clicks
Average Position	Average ranking in Google Search results	Average ranking in Google Search results
CTR	Click-through rate. Clicks/Impressions for Google Search clicks.	Click-through rate. Clicks/Impressions for both Google Ads and Google Search clicks.
Keyword	Applies to the key terms used in the written content of the website pages. These terms are the most significant keywords and their variants that Google found when crawling your site. When reviewed along with the Search queries report and your site's listing in actual search results for your targeted keywords, it provides insight into how Google is interpreting the content of your site.	In paid-search or Google Ads reports, describes a paid keyword from a search-engine-results page. In the organic-search reports, describes the actual query string a user entered in a web search.
Query	The actual query a user entered in Google search.	Only used in the Search Console reports. Applies to the actual query a user entered in Google search.

Discrepancies in Data between GA and GSC

Context	Search Console	Analytics

Landing-page URLs that redirect	Search Console reports the Canonical URL for a landing page, even when the click was to a non-Canonical landing page.
	
	If
	
	`www.example.com/amp`
	
	has a canonical URL of
	
	`www.example.com`
	
	Search Console reports search metrics for
	
	`www.example.com`
		Analytics reports the URL that results from the redirect, for example:
		
		`www.example.com`

Page has no Analytics tracking code	Data for the page appears in Search Console.	Data for the page does not appear in Analytics.
Number of URLs recorded per site per day	Search Console records up to 1000 URLs for landing pages.	Analytics does not observe the 1000-URL limit, and can include more landing pages.
Analytics property tracks multiple domains	Search Console can link to a single domain.	If an Analytics property collects data for multiple domains, the Search Console reports have data for only the single linked domain.
Time zones vary	Search Console timestamps data according to Pacific Daylight Time.	Analytics timestamps data in each view according to the time zone identified in the view settings.
JavaScript not enabled in browsers	Search Console collects data regardless of whether JavaScript is enabled.	Analytics collects data only when JavaScript is enabled. Users can opt out of data collection by implementing a browser add-on.

At the time of this book's publishing, Google's data retention policy for GSC data is 16 months, and Bing's Webmaster Tools data retention policy is 6 months. If you want to maintain a history of the data provided by these tools, you should plan on regularly downloading and saving your data. Many third-party analytics, marketing, and business intelligence platforms provide the ability to pull GSC and BWT data down via API data source connectors.

First Party Data and The Cookie-Less Web

As a result of the increased emphasis on user data privacy on the web, digital marketers are working in an increasingly cookie-less environment, whereby first-party data is king and alternative methods for Google's "Privacy Sandbox" (https://privacysandbox.com/) is Google's solution to the deprecation of third-party cookies and migrating web technologies towards a more privacy-focused web. Google's migration to Google Analytics 4 in July of 2023 represents a shift from the "pageview" metric in Universal Analytics to an event-based model, moving away from "bounce rates" to engagement rates. As a result of the cookie-less migration, much of this new reporting will be reliant upon model-based estimations of data, versus actual data.

Much has yet to be seen in terms of how digital marketers will adapt to these types of shifts within the major analytics tools and platforms. Topics of data privacy, regulation, and related legal issues are covered further in Chapter 12.

Conclusion

For SEO professionals, effective tracking, measurement, and analysis are critical to the ability to effectively implement and optimize your SEO efforts. Putting the right analytics tools in place can provide tremendous

insight to empower this process, and ensure that you continue to build a strong foundation for SEO success.

CHAPTER SEVEN

Google Algorithm Updates and Manual Actions/Penalties

Google tunes and tweaks their algorithm on a daily basis, and they periodically release larger algorithm updates. In addition, they actively review their results to find sites that are violating their guidelines (https://bit.ly/webmaster_best_practices), and those sites may be subjected to ranking penalties. All of these measures are designed to help them improve the quality of their overall search results.

Sometimes these algorithm changes and penalties can have a major impact on your organic traffic. Significant decreases in the search engine traffic to your website can be devastating to a business. As shown in Figure 7-2, sometimes these drops in traffic can be quite large.

Figure 7-1. Major loss in traffic

If the business shown in Figure 7-2 generates most or all of its revenue from organic search traffic, this would represent a crippling blow. This type of loss of revenue can mean laying off employees, or even closing the business.

For that reason, you need to have a working understanding of how the Google ecosystem works, how Google recommends that you operate your website, and the various scenarios which can lead to visibility and traffic losses. Otherwise you may be impacted by Google updates or penalties and it may seem like this happened for reasons beyond your control. However, with the right understanding of what Google is trying to do overall with their algo updates and penalties you can significantly reduce your exposure to them and potentially set yourself up to avoid the penalties and benefit from the updates. If you have already suffered such a traffic loss or become impacted by an update or penalty, it is important to understand what the cause is and what you need to do to recover.

Google Algorithm Updates

Google's updates to its various search algorithms take many different forms, including changes to search functionality, updates to search result composition and layout, changes in various aspects of relevance and ranking algorithms, and daily testing and bug fixes. In this section we'll review the types of changes that Google makes and their impact on the search results that users ultimately engage with.

BERT

On October 25, 2019 Google announced the existence of a new algorithm called BERT (https://www.blog.google/products/search/search-language-understanding-bert/). BERT is short for Bidirectional Encoder Representations from Transformers, and it is a neural network-based technique for natural language processing (NLP). This is what Google said about the impact of BERT: "BERT will help Search better understand one in 10 searches in the U.S. in English, and we'll bring this to more languages and locales over time."

Prior to BERT, when Google's algorithms were trying to understand the meaning of a word or phrase it could only consider nearby text that came before that word or phrase, OR after it. Essentially, it was unidirectional. With BERT, Google is now able to analyze the text before AND after the word or phrase to understand its meaning. Figure 7-2 shows an example of a query impacted by Google.

2019 brazil traveler to usa need a visa

BEFORE

AFTER

9:00

google.com

wp Washington Post › 2019/03/21
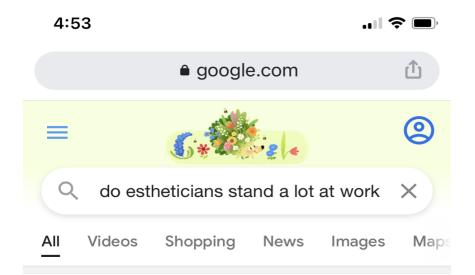
U.S. citizens can travel to Brazil without the red tape of a visa ...

Mar 21, 2019 · Starting on June 17, you can go to Brazil without a visa and ... Australia, Japan and Canada will no longer need a visa to ... washingtonpost.com; © 1996-2019 The Washington Post ...

9:00

google.com

USEmbassy.gov › br › Visas

Tourism & Visitor | U.S. Embassy & Consulates in Brazil

In general, tourists traveling to the United States require valid B-2 visas. That is unless they are eligible to travel visa ...

Figure 7-2. Sample Search Query Before and After BERT
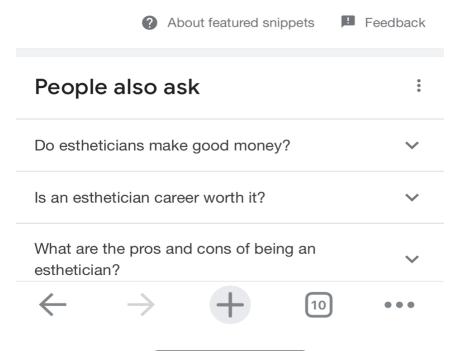
As illustrated in the figure, prior to BERT, Google did not properly understand the query and hence did not answer the user's question. In the result after BERT's impact content surfaces that looks like Google fully understands the context of the question. As with all Google's algorithms they evolve and improve over time, and the results continue to improve as shown in Figure 7-3.

4:53

🔒 google.com

☰　🌸（Google Doodle）　👤

🔍　do estheticians stand a lot at work　✕

All　Videos　Shopping　News　Images　Maps

When it comes to the question of **do estheticians stand a lot at work**, the answer is absolutely, **estheticians** indeed **stand a lot at work**.

nathanieltower.com › do-esthetician...　⋮

Do estheticians stand a lot at work - Nathaniel Tower

❓ About featured snippets　　🚩 Feedback

People also ask　⋮

Do estheticians make good money?　⌄

Is an esthetician career worth it?　⌄

What are the pros and cons of being an esthetician?　⌄

←　→　＋　🔟　•••

Figure 7-3. BERT Continues to Evolve and Improve

In the initial rollout BERT was only applied to US language queries, and Google indicated that it had an impact on 10% of those queries. On December 9, 2019 they announced that BERT was further rolled out to 70 languages. https://searchengineland.com/bert-is-rolling-out-to-google-search-in-over-70-languages-326146

In addition to BERT, Google published a paper on a new algorithm called SMITH. This is an algorithm that could be the next step beyond BERT. What SMITH could add is improved capabilities understanding longer passages within long documents the same way that BERT understands words or phrases. As of November 2021 it was not clear whether the SMITH algorithm had not been rolled out into Google Search, but it illustrates that continuing investigation by Google into improved natural language processing is ongoing.

Passages and Subtopics

On October 15, 2020 Google announced their intention to release two new search algorithms (https://www.blog.google/products/search/search-on/). The first of these to be release was an algorithm that would enable them to divide their search results into topics and subtopics. The genesis of this was Google's recognition that in many cases broad user queries get followed rapidly by additional queries designed to further refine what the user is looking for.

Figure 7-4 shows an illustration of how this might work. For example, if a user searches on "Home Exercise Equipment" Google may show some initial results for that query but also add subsections for "Affordable Exercise Equipment" and "Small Space Exercise Equipment," as these are popular follow-on queries for users.

Google | Home Exercise Equipment | x | 🎤 🔍

Home Exercise Equipment A1

Home Exercise Equipment A2

Affordable Exercise Equipment

Small Space Exercise Equipment

Figure 7-4. Illustration of the Subtopics Algorithm

Google's Danny Sullivan confirmed in January 2021 that the subtopics algorithm had been rolled out by Google in Mid-November 2020.
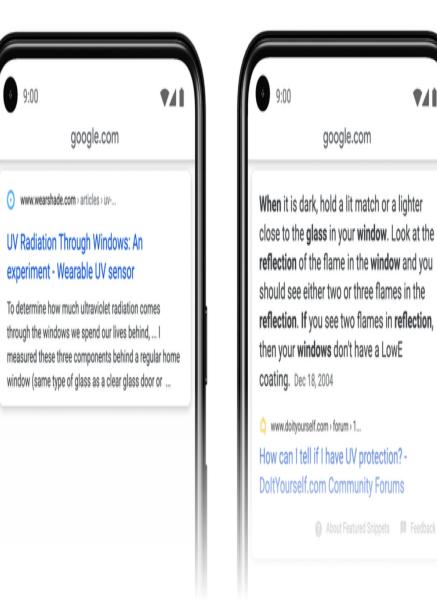
Another one of the announced algorithms was one that would enable them to identify and "index" specific passages within a web page separately from the content of the rest of the page. The purpose of this update was to allow them to provide answers to very specific user questions. Figure 7-5 shows the example of such a query that Google shared in their announcement.

how can I determine if my house windows are UV glass

BEFORE

AFTER

9:00

google.com

www.wearshade.com › articles › uv-...

UV Radiation Through Windows: An experiment - Wearable UV sensor

To determine how much ultraviolet radiation comes through the windows we spend our lives behind, ... I measured these three components behind a regular home window (same type of glass as a clear glass door or ...

9:00

google.com

When it is dark, hold a lit match or a lighter close to the **glass** in your **window**. Look at the **reflection** of the flame in the **window** and you should see either two or three flames in the **reflection. If** you see two flames in **reflection**, then your **windows** don't have a LowE coating. Dec 18, 2004

www.doityourself.com › forum › 1...

How can I tell if I have UV protection? - DoItYourself.com Community Forums

About Featured Snippets   Feedback

Figure 7-5. Example of the Passages Algorithm in Action

The reason that this algorithm is important is that many user needs are highly specific. While the answers to these queries may exist in various places on the web, many of these answers may be buried inside other content whose general relevance may not align well with the specific user question.

With this update, Google can start to recognize specific passages within a larger document that are relevant to such a specific question. Google's Danny Sullivan (aka @SearchLiaison on Twitter) announced that the initial release of the Passages algorithm had been rolled out on February 11, 2011.
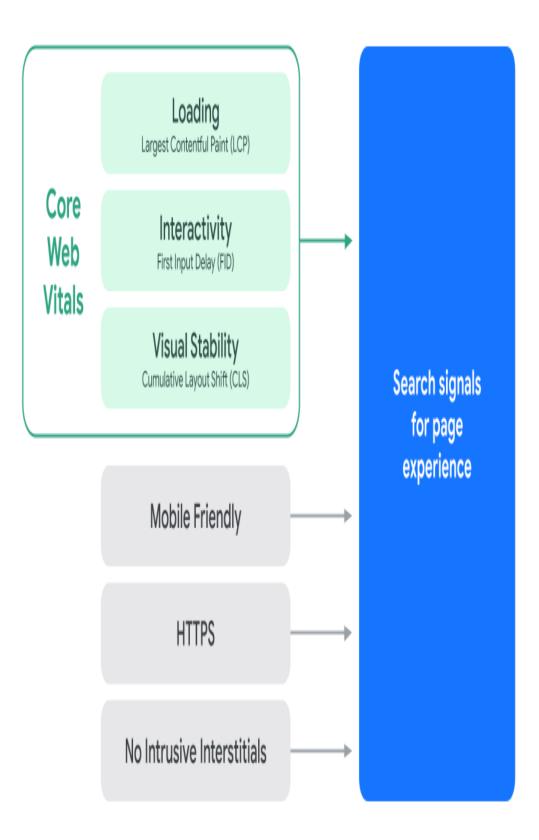
Page Experience and Core Web Vitals

On May 28, 2020 Google announced their intention to begin using a new signal called Page Experience (https://developers.google.com/search/blog/2020/05/evaluating-page-experience). This was then followed by multiple clarifications on the timing of the roll out for the new signal. As of August 2021 the status of the Page Experience rollout was that it began rolling out in mid-June 2021 and was expected to complete by the end of August 2021 (https://developers.google.com/search/blog/2021/04/more-details-page-experience).

The Page Experience signal is actually a collection of many pre-existing signals all related to whether or not your site offers a good experience to users. Figure 7-6 provides Google's visual representation of what the signals are and how they come together into one signal.

Core Web Vitals

Loading
Largest Contentful Paint (LCP)

Interactivity
First Input Delay (FID)

Visual Stability
Cumulative Layout Shift (CLS)

Mobile Friendly

HTTPS

No Intrusive Interstitials

Search signals for page experience

Figure 7-6. Google's Page Experience signal

A more detailed description of each of the components is as follows:

- Core Web Vitals: These are metrics related to page speed and visual stability. This comprises three components:

    - Largest Contentful Paint (LCP) – A measure of how long it takes for the main content elements of the page to draw. Web.dev's precise definition for this is: "the render time of the largest image or text block visible within the viewport, relative to when the page first started loading." (https://web.dev/lcp/) The emphasis on viewport is an important one as this means that the true focus is on rendering of the content "above the fold". Google suggests that users try to have this occur in 2.5 seconds or less.

    - First Input Delay (FID) – Web.dev defines FID as follows: "FID measures the time from when a user first interacts with a page (i.e. when they click a link, tap on a button, or use a custom, JavaScript-powered control) to the time when the browser is actually able to begin processing event handlers in response to that interaction." (https://web.dev/fid/). This is essentially a measure of how quickly things begin to happen after a user clicks on something on your web page. Google suggests that you target a value here of 0.1 seconds or less.

    - Cumulative Layout Shift (CLS) – A measure of visual stability. The reason this is important is that there are many sites that begin painting content and then the page content appears to jump around as it continues to draw. The reason that this is really bad is that the user may attempt to click on something during the process only to have the page jump around right at that moment and they end up actually clicking on something that they didn't

intend. Web.dev defines CLS as: "the sum total of all individual layout shift scores for every unexpected layout shift that occurs during the entire lifespan of the page" (https://web.dev/cls/)

- Mobile Friendly: Whether or not your site offers a good experience on mobile devices. You can test this with Google's Mobile Friendly Test Tool (https://search.google.com/test/mobile-friendly).

- Secure Browsing: Whether or not you have implemented HTTPS/TLS for your web site. You can learn more about how to do that by reading Google's page on how to Secure Your Site with HTTPS (https://support.google.com/chrome/answer/95617). You can also learn how to see if a site is secure by visiting the Chrome "Check if a site's connection is secure" page. (https://support.google.com/chrome/answer/95617).

- No Intrusive Interstitials: The absence of interstitials that block user's access to the information they came for, particularly on initial page load.

The reason that Google has organized all these signals into one larger score in the overall Google algorithm is that it makes the concept of Page Experience as a ranking factor much simpler for them to manage. The overall weight of Page Experience can be treated as one signal, and how the relative weighting of the individual components works can be decided in isolation from the main algorithm.

In addition, when a new page experience related signal is decided upon by Google this can easily be added into the Page Experience signal without having to tinker with the larger algorithm.

While Page Experience is important, remember that content relevance and quality are always the most important signals. To illustrate, your page about tadpoles will not begin ranking for user search queries about cooking pots just because it's fast. Similarly, a strong Page Experience score will not help your poor quality content rank, even if it is highly relevant.

However, there are many instances where queries are highly competitive, with many potential pages that offer high relevance and high quality to address what the user is looking for. In such cases, the Page Experience signal can play a key role in helping you rank just a bit higher than your competition.

Link Spam Update

Another significant update released by Google in 2021 was focused on link spam. The simply named Link Spam Update began rolling out on July 26, 2021 and completed its rollout on August 24, 2021. The nature of this update was outlined by Google in a post called "A reminder on qualifying links and our link spam update" (https://developers.google.com/search/blog/2021/07/link-tagging-and-link-spam-update)

While the post does not specifically discuss what the link spam update addressed, the start of the post does have a focused discussion on problems with affiliate links and guest posting. This includes a reminder of how important it is to use link tags such as NoFollow, Sponsored, and UGC where appropriate.

That does not mean that other aspects of link spam weren't potentially addressed as well but does suggest that these areas were the core focus.

What it does tell us is that fighting link spam remains a real issue for Google, and while it's clear that they have made a lot of progress over the years, they still have room for improvement.

Broad Core Algorithm Updates

Starting in March of 2018 Google began to announce what it calls Broad Core Algorithm Updates (BCAU). Figure 7-7 shows how Google's Danny Sullivan summarized the March 28, 2018 update on Twitter:

Barry Schwartz ✔ @rustybrick · Mar 12, 2018

Hey @JohnMu or @dannysullivan any specific update you want to confirm or deny or something? details at

> Google Algorithm Update Over The Weekend - ...
> There may have been an algorithm update with the Google search rankings over the weekend, ...
> 🔗 seroundtable.com

Danny Sullivan ✔
@dannysullivan

Yes, see here:

> G **Google SearchLiaison** ✔ @searchliaison
>
> Each day, Google usually releases one or more changes designed to improve our results. Some are focused around specific improvements. Some are broad changes. Last week, we released a broad core algorithm update. We do these routinely several times per year....

12:58 PM · Mar 12, 2018

Figure 7-7. Danny Sullivan Tweet About the March 2018 Algo Update

Since that date Google has been regularly announcing these larger updates, and as Sullivan indicated, these happen several times per year. It's also important to note that these confirmed updates by Google are ones that they deem significant enough to confirm, but Google in fact makes many more updates that they choose to not make any comment on. In reverse order, from newest to oldest, the dates of these core algorithm updates have been:

- November 3, 2021
- July 1, 2021
- June 2, 2021
- December 3, 2020
- May 4, 2020
- January 13, 2020
- September 24, 2019
- June 3, 2019
- March 12, 2019
- August 1, 2018
- April 17, 2018
- March 7, 2018

In addition to all of these confirmed updates, the industry has noted many other dates where algorithm changes by Google appear to have had a larger impact. These unconfirmed updates are sometimes quite significant in their impact with many websites gaining or losing traffic. Further, Google makes tweaks to their algorithms on a daily basis. In July of 2019 Google's Danny Sullivan noted that they had made more than 3,200 changes to their

algorithm in the past year (https://www.blog.google/products/search/how-we-keep-google-search-relevant-and-useful/).

Functionality Changes

Google periodically makes tweaks to the way that search works. Some of these are announced as well. Examples as of the writing of this book include:

Passage Indexing Update – February 10, 2021

This update marked the initial roll out of Passages functionality into the core algorithm. This is discussed in more detail in the Passages and Subtopics section of this chapter.

Featured Snippet De-duping — January 22, 2020

Prior to this update pages shown in Featured Snippets would also be shown as a regular search result further down in the SERPs. This approach was a result of the Featured Snippet algorithm being a separate one from the main Google ranking algo. With this update the Featured Snippet algorithm was effectively fully integrated. As a result, Google stopped showing the regular search result for pages highlighted in the Featured Snippet.

You can read more about this update here: https://www.seroundtable.com/google-featured-snippet-deduplication-change-summary-28921.xhtml.

BERT Update — October 22, 2019

This update expanded the use of BERT to 70 different languages. This is discussed in more detail in the BERT section of this chapter.

Site Diversity Update — June 6, 2019

Prior to this update Google had many instances of search results where many listings would come from the same domain. In the Site Diversity Update the frequency of these occurrences was reduced.

You can read more about this update here: https://moz.com/blog/did-googles-site-diversity-update-live-up-to-its-promise

Chrome Non-Secure Site Warnings – July 24, 2018

While not an update to the Google search algorithms, the release of Chrome 68 is when Google began labelling all sites that were not running on HTTPS as Not Secure. For sites that had not converted as of that date this was an action that likely impacted their traffic.

Mobile Page Speed Update – July 9, 2019

With this release page speed officially became a ranking factor for mobile results. Google stated then, and continues to state, that page speed only impacts the slowest of sites.

Video Carousels Update – June 14, 2018

Google moved all videos appearing in the search results into video carousels causing significant changes in clickthrough rates in the SERPs.

Search Snippet Length Adjustments – May 13, 2018

Google formally ended tests that it had been running on showing longer snippets for some queries in the search results.

Mobile-first Index Roll Out – March 26, 2018

The formal process of moving sites into the mobile-first index began on this date (https://developers.google.com/search/blog/2018/03/rolling-out-mobile-first-indexing). The move towards mobile-first indexing was first announced on November 4, 2016 (https://developers.google.com/search/blog/2016/11/mobile-first-indexing) and Google had largely been testing the potential impact since then and up until this announcement. Even after this announcement the continuing move of sites into the mobile-first index took place at a slow pace to minimize disruption to the search results.

Zero Results SERP Test – March 14, 2018

For some Knowledge Card results Google tested zero result SERPs combined with a "Show all results" button. An example of one of these results is shown in Figure 7-8. Per Danny Sullivan, this test was cancelled on March 20, 2018, likely because Google determined that this format was not appreciated by users.

10:51 AM

Wednesday, March 14, 2018 (PDT)

Time in Seattle, WA

Show all results ›

Figure 7-8. Example of a Google Zero Result SERP

You can read more about this update here: https://moz.com/blog/zero-result-serps.

Google Bug Fixes

Google search is a large and complex ecosystem, and it's inevitable that bugs will show up in the system from time to time. For reference, here is a summary of the Google confirmed bug fixes from 2019 through the start of 2021:

- September 29, 2020 to October 12, 2020: Google confirmed that an indexing bug was resolved.

- August 10, 2020: Google confirmed that an indexing bug was resolved.

- June 22, 2020: Google confirmed that an indexing bug was resolved.

- May 23, 2019: Google confirmed that an indexing bug was resolved.

- April 5, 2019: Google confirmed that an indexing bug was resolved.

It is also highly probable that there are large quantities of bug fixes made by Google that we never hear about.

Google Webmaster Guidelines

If you're the owner/publisher of a website who wants to grow your traffic from Google it's valuable to develop a strong understanding of Google's Webmaster Guidelines (https://bit.ly/webmaster_best_practices). These detail the principles that Google wants webmasters to follow with their websites. While Google can't require you to follow these guidelines they can choose to give poorer rankings to websites that don't. The basic principles that Google wants webmasters to follow are:

Make pages primarily for users, not for search engines

> This is a critical aspect for the web presence of any business. Knowing what your target users want, how they search, and how to create a site that presents that in an understandable and engaging way is just good business, and it's good for ranking in Google as well.

Don't deceive your users

> Sadly, this one is here because there are many web sites that use bait and switch tactics to draw users into content and experiences which aren't what they expected. For example, sites that have problems with Cumulative Layout Shift (discussed in the Page Experience section of this chapter) may cause users to accidentally click on the wrong part of a page on the site, creating a very poor user experience.

Avoid tricks intended to improve search engine rankings. A good rule of thumb is to ask yourself whether you'd feel comfortable explaining what you've done to a website that competes with you, or to a Google employee. Another useful test is to ask, "Does this help my users? Would I do this if search engines didn't exist?"

> Take the final sentence of this principle to heart. At one level, it may sound naïve, but when you realize that Google actively tunes its algorithms to find the sites that do the best job of serving users it starts to make more sense. Google is tuning all of its algorithms to find the best user experiences and as a result focusing your efforts on creating great value for users is strongly aligned with maximizing your chances of ranking in Google.

Think about what makes your website unique, valuable, or engaging. Make your website stand out from others in your field .

> A focus on users is necessary but not sufficient. You must also strive to create a site that stands out, just as you seek to have a business that stands out from competition. Otherwise there will be nothing about it

that compels users to want to see your website and likewise provides Google with little reason to rank it highly in the search results.

In addition to these basic principles, Google also discusses a number of specific guidelines. These are divided into practices to avoid as well as a couple of good practices to follow.

Practices to Avoid

Automatically generated content

Here Google is targeting pages that are artificially generated for purposes of attracting search traffic, and that add no practical unique value. Of course, if you run a retail site you may be using your ecommerce platform to automatically generate pages representing your database of products, and that is not Google's concern here. This is more targeted at machine-generated (aka mad-libbed) content that makes little sense to users.

Participating in link schemes

Since links to your site remain a core part of the Google algorithm there are many parties out there offering ways to cheaply, and artificially, generate links to your site. As discussed in the Quality Links section of this chapter, focus your efforts instead in attracting links that represent legitimate citations of your site.

Creating pages with little or no original content

This can take many forms, such as pages that are automatically generated, pages with little user value or purpose but exist just to get someone to click on an affiliate link, content scraped from other sites or doorway pages.

Cloaking

Google defines this as "the practice of presenting different content or URLs to human users and search engines." The reason that this is an issue is that some websites were structured to show Google a rich informational experience that Google might choose to rank, but when the user arrived at the site they would get something entirely different.

Sneaky redirects

This is the practice of using redirects to send users to a different page than what gets shown to Googlebot. As with cloaking the concern is that users may get sent to content that does not match up with what they expected when they clicked on a link in a Google search result.

Hidden text or links

These are spammy tactics that date back to the beginning of search engines where content is rendered on a page in a way that it's not visible, such as implementing white text on a white background, or using CSS to position it well off the page. With links, a common spam tactic was to include a link to a page but implementing a link on only one character such as a hyphen.

Doorway pages

These consist of pages that were created solely to attract search engine traffic and not for the purpose of creating a great user experience. In practice, these often are created in high volume and are not well integrated into the rest of the web site. They also may be designed to target lots of highly similar, though not exactly identical, search phrases.

Scraped content

Taking content from third party sites ("scraping" it) and republishing it on your own site is not only a copyright violation, but Google frowns on it as well. Making minor modifications such as using synonyms is not sufficient either. If you're going to quote something from another site

you must provide a citation to that site and add your own unique value to it.

Participating in affiliate programs without adding sufficient value

Historically Google had a lot of problems with sites that generated all of their revenue from affiliate programs finding ways to rank poor quality content and still rank in the search results. There is nothing wrong with making some, or all, of your revenue from affiliate programs. Figure 7-9 shows what Google had to say about one affiliate site indicating that being an affiliate site is not a problem by itself. However, if you offer little content of value to users Google will not want to rank that site.

Gary 鯨理 / 경리 Illyes ✔

@methode

Replying to @rishil

have you seen what @thepointsguy is doing? They create amazing content & they have sponsored links, but the focus is on the user

3:49 AM · Mar 14, 2017 from Zurich, Switzerland · Twitter for iPhone

Figure 7-9. Google's Gary Illyes Praises an Affiliate Site's Content

Loading pages with irrelevant keywords

Also called "keyword stuffing," loading your page with irrelevant or overly repeated uses of words creates a poor user experience and is also seen as spammy behavior by Google.

Creating pages with malicious behavior , such as phishing or installing viruses, trojans, or other badware

The reasons why Google would not want to serve up these pages in the search results is obvious, but these are not always the result of action by the publisher of the website. Sites can get hacked and it pays to be vigilant in maintaining the security of your site and regularly checking to see if it has been hacked.

Abusing structured data markup

Structured data provides you with opportunities to enhance the appearance of your listing in Google's search results, but there is the potential for abuse here too.

Sending automated queries to Google

This is the practice of using automated tools to send large quantities of queries to Google. This type of activity is often used for rank tracking purposes and Google does not care for it as it uses up their resources without any return benefit for them. Note though that many tools do large scale rank tracking, such as Brightedge, Searchmetrics, SEMrush, seoClarity, Conductor, and others. Using one or more of these tools can be a highly valuable component of your SEO program so just ensure that you don't overdo it.

Good Hygiene Practices to Follow

This list is comparatively short and focuses on two areas that represent best practices site hygiene.

Monitoring your site for hacking and removing hacked content as soon as it appears

This is unfortunately more common than might expect. Hackers create programs to comb the web looking for sites that have security vulnerabilities and then using those to inject their code into your web pages, often to insert invisible links to their web pages.

One of the best defenses you can implement to limit your risk here is to always keep your software platform up to date. For example, if you use Wordpress, always install the latest updates very soon after they become available to you. This would include any plugins as well.

Preventing and removing user-generated spam on your site

Any site that allows users to contribute content in any form has a high risk of having spammy content put there by users. For example, if you allow comments on the content you publish, or host forums on your site. Some bad actors can come in and manually put in spammy content and even worse actors implement programs that crawl the web looking for places to place comments or posts on sites.

Some of the best practices here include requiring moderation of all comments or posts or reviewing all comments or posts shortly after they have been put up. There are gradations of this too such as requiring moderation of the first comment or post by any user, but then letting them contribute additional content without moderation thereafter. However, you should still plan to review those contributions once they're posted.

User generated spam can also occur on freehost sites – places where sites can be put up without any financial commitment. If you operate a freehost platform you will need to implement similar procedures to ensure that you are not getting spammy content implemented on your site.

Do take the time to read through the Google Webmaster Guidelines (https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines). Anyone who begins to proactively invest in increasing their organic search presence should understand them and take steps to ensure that their organization does not cross the lines.

Quality Content

Since we, as publishers of website content, want traffic from Google, it becomes our task to provide high quality content. This requires an understanding of our target audience, how and what they search for, and then providing high quality content wrapped in a great user experience so they can find what they want quickly.

However, as you might expect, creating high quality content is not always that easy and many parties attempt to take shortcuts that can potentially result in low quality content, or even spam content, showing up in the search results. To combat this there are many things that Google does and looks for to ensure that it minimizes the presence of poor-quality content in its SERPs.

A decade ago, Google took a big step forward when it released the Panda algorithm on February 24, 2012. In their announcement of the release (http://bit.ly/more_high-quality), Google said the following:

Many of the changes we make are so subtle that very few people notice them. But in the last day or so we launched a pretty big algorithmic improvement to our ranking—a change that noticeably impacts 11.8% of our queries—and we wanted to let people know what's going on. This update is designed to reduce rankings for low-quality sites—sites which are low-value add for users, copy content from other websites, or sites that are just not very useful. At the same time, it will provide better rankings for high-quality sites—sites with original content and information such as research, in-depth reports, thoughtful analysis, and so on.

The big change that Panda brought to the Google landscape is that it brought Google's capabilities for evaluating content quality to another

level. Part of this involved downgrading sites that were publishing low quality content in high volume to drive large quantities of search traffic. But, over time Panda was adapted to address issues of content quality at a much broader level.

Initially Panda was an algorithm that ran separately from the main Google algorithm, but in January of 2016 Google confirmed that they had fully integrated Panda into their main algorithm. Content quality remains a critical component of Google's algorithms.

Content That Google Doesn't Like

Some of the key types of content that that Google considers to be poor are as follows:

Thin content

> As you might expect, this is defined as pages with very little content. Examples might be user profile pages on forum sites with very little information filled in, or an ecommerce site with millions of products, but very little information provided about each one.

Unoriginal content

> These may be scraped pages, or pages that are only slightly rewritten, and Google can detect them relatively easily. Sites with even a small number of these types of pages can be negatively impacted by Google algorithms.
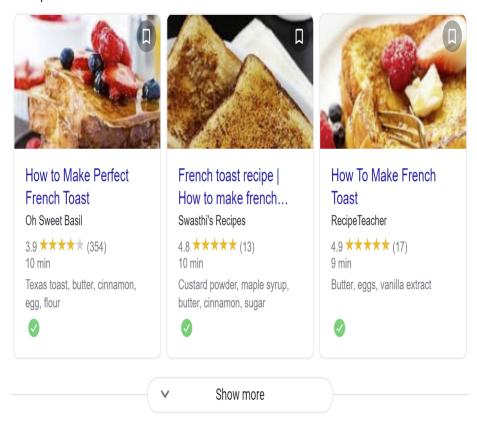
Nondifferentiated content

> Even if you create all original articles, this may not be enough. If every page on your site covers topics that others have written about hundreds or thousands of times before, then you really have nothing new to add to the Web with your site. Consider, for example, the number of articles in the Google index about making French toast, as shown in Figure 7-10. There are over 5,000 pages in the Google index that include the phrase

how to make french toast in their title. From Google's perspective, it doesn't need more web pages on that topic.

Google

intitle:"how to make french toast"

All · Videos · Books · Images · Shopping · More   Settings   Tools

About 5,180 results (0.93 seconds)

Recipes

How to Make Perfect French Toast
Oh Sweet Basil
3.9 ★★★★☆ (354)
10 min
Texas toast, butter, cinnamon, egg, flour

French toast recipe | How to make french...
Swasthi's Recipes
4.8 ★★★★★ (13)
10 min
Custard powder, maple syrup, butter, cinnamon, sugar

How To Make French Toast
RecipeTeacher
4.9 ★★★★★ (17)
9 min
Butter, eggs, vanilla extract

Show more

www.youtube.com › watch

How to Make French Toast!! Classic Quick and Easy Recipe - YouTube

so what I'm gonna do is just kind of mix up my French toast. my batter mixture a little bit here make sure that ...

5:11

Sep 10, 2016 · Uploaded by Crouton Crackerjacks

Poor-quality content

This is content that is inaccurate or poorly assembled. In many cases, this may be hard to detect, but one indicator is content that includes poor grammar or a lot of spelling mistakes. Google could also potentially use fact checking as another way to determine poor-quality content.

Curated content

Sites that have large numbers of pages with lists of curated links do get hit by Google algorithms. Content curation is not inherently bad, but if you are going to do it, it's important to incorporate a significant amount of thoughtful commentary and analysis. Pages that simply include lots of links will not do well, nor will pages that include links and only a small amount of unique text.

Thin slicing

This used to be a popular tactic for *content farms.* Imagine you wanted to publish content on the topic of schools with nursing programs. Content farm sites would publish many articles on essentially the same topic. An example would be creating highly similar articles with titles such as: "nursing schools," "nursing school," "nursing colleges," "nursing universities," "nursing education," and so forth. There is no need for all of those different articles, as they will have no material differentiation from each other.

Database-generated content

The practice of using a database to generate web pages is not inherently bad, but many companies were doing it to an excessive scale. This can lead to lots of thin-content pages or poor-quality pages, which Google does not care for. Note that ecommerce platforms essentially generate

content from a database, and that's OK as long as you also work to get strong product descriptions and other information on those pages.

Importance of Diverse Content

Diversity is important to overall search quality for Google. One simple way to illustrate this is with the search query *Jaguar.* This word can refer to an animal, a car, a guitar, an operating system, or even an NFL team. Normal ranking signals might suggest the results shown in Figure 7-11.

Figure 7-11. Normal ranking signals may show these results for "Jaguar"

Note that the search results at the top all focus on the car, which may be what the basic ranking signals suggest the searcher is looking for. However, if the searcher is looking for information on the animal, those results are

pushed down a bit. As a result, Google may use other signals to decide to alter the results to look more like those shown in Figure 7-12.

Figure 7-12. "Query Deserves Diversity" may alter the results for "Jaguar"

In this version of the results, one of the animal-related results has been inserted into the second position. Google makes these types of adjustments to the SERPs using a concept known as *Query Deserves Diversity*.

Google makes these adjustments by measuring user interaction with the search results to determine what ordering of the results provides the highest levels of user satisfaction. For example, even if traditional ranking signals would put another page for the car next, it might make sense for the next result to be about the animal as the result might be a higher percentage of satisfied users.

Role of Authority in Ranking Content

Consider again the search query *how to make French toast* we showed in While Google has plenty of results on the topic, there are, of course, some sites that rank highly for this search query. How is their rank determined?

Very high-authority sites are likely to do fine when publishing content on a topic that is already well covered on the Internet. There are a few possible reasons why this is the case:

- Reputation and authority are a big factor. For example, if the New York Times Lifestyle section posted a new article on how to make French toast, even though it is not particularly unique, readers might respond well to it anyway. User interaction signals with the search result for that content would probably be quite strong, simply because of the site's reputation.

- High-authority sites probably got to be that way because they don't engage in much of the behavior that Google advises webmasters to avoid. Chances are that you won't find a lot of thin content, "me too" content, thin slicing, or any of the issues that are triggers for Google's algorithms.

- Google may simply be applying looser criteria to a high-authority site than it does to other sites.

Exactly what factors allow higher-authority sites to have more leeway is not clear. Is it that Google is measuring user interaction with the content, the quality of the content itself, the authority of the publisher, or some combination of these factors? There are probably elements of all three in what Google does.

Impact of Weak Content on Rankings

Weak content on even one single section of a larger site can cause Google to lower the rankings for the whole site. This is true even if the content in question makes up less than 20% of the pages for the site. As Figure 7-13 shows, this may not be a problem if the rest of your site content has really strong content, but it's best not to take the chance if you have known weak pages and it's worth the effort to address them.

Danny Sullivan ✔
@dannysullivan

Replying to @dannysullivan and @schachin

Want to do better with a broad change? Have great content. Yeah, the same boring answer. But if you want a better idea of what we consider great content, read our raters guidelines. That's like almost 200 pages of things to consider:

static.googleusercontent.com/media/www.goog... ✅

5:39 PM · Aug 1, 2018 · Twitter Web Client

Figure 7-13. Impact of weak content on rankings

Improving Weak Content

When addressing thin content it's best to dig deep and take on hard questions about how you can build a site full of fantastic content that gets lots of user interaction and engagement. Highly differentiated content that people really want, enjoy, share, and link to is what you want to create on your site. There is a science to creating content that people will engage with.

We know that picking engaging titles for the content is important, and that including compelling images matters too. Make a point of studying how to create engaging content that people will love and apply those principles to every page you create. In addition, measure the engagement you get, test different methods, and improve your ability to produce great content over time.

Ways to address weak pages

As you examine your site, a big part of your focus should be addressing its weak pages. They may come in the form of an entire section of weak content, or a number of pages interspersed among the higher-quality content on your site. Once you have identified those pages, there are a few different paths you can take to address the problems you find:

- Improve the content. This may involve rewriting the content on the page, and making it more compelling to users who visit.

- Add the noindex meta tag to the page (you can read about how to do this in Chapter 6). This will tell Google to not include these pages in its index, and thus will take them out of the Panda equation.

- Delete the pages altogether, and 301-redirect visitors to other pages on your site. Use this option only if there are quality pages that are relevant to the deleted ones.

- Delete the pages and return a 410 HTTP status code when someone tries to visit the deleted page. This tells the search engine that the pages have been removed from your site.

- Use the URL removal tool (http://bit.ly/remove_content) to take the page out of Google's index. This should be done with great care. You don't want to accidentally delete other quality pages from the Google index!

Quality Links

To understand how Google uses links, we need only review Larry Page and Sergey Brin's original thesis, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (http://infolab.stanford.edu/~backrub/google.xhtml). At the beginning of the thesis is this paragraph:

The citation (link) graph of the web is an important resource that has largely gone unused in existing web search engines. We have created maps containing as many as 518 million of these hyperlinks, a significant sample of the total. These maps allow rapid calculation of a web page's "PageRank," an objective measure of its citation importance that corresponds well with people's subjective idea of importance. Because of this correspondence, PageRank is an excellent way to prioritize the results of web keyword searches.

The concept of a *citation* is critical. Consider the example of an academic research paper, which might include citations similar to those shown in Figure 7-14

Franklin, K. Anabaptists: A Complete History [Book].
New York: Siegfried Shaw. 2002 Aug - [Cited 2004 May 12]

James, R. Time and Space and Me. [Book].
San Franciso: Abington Krane. 1996 Sep - [Cited 2010 Aug 17]

Landry, P. Automotive Engineering III. [Book].
Boston: Helmsworth Claw. 1984 Jan - [Cited 2005 Feb 1]

Figure 7-14. Academic citations

The paper's author uses the citation list to acknowledge major sources he referenced as he wrote the paper. If you did a study of all the papers on a given topic area, you could fairly easily identify the most important ones, because they would have the most citations (votes) by other papers.

To understand why links have value as a signal for search engines one need only to consider what they represent. When someone links to your website,

they're offering users the opportunity to leave their web site and go to yours. Generally speaking, most website publishers work hard to bring as many people to their website as they can.

Then they want those people to complete some action of value on their site, such as buying something, viewing ads, visiting many pages so they see many ads, or click on ads. For some sites where expressing a strong position on a debated topic matter is the goal it may simply be to get the user to read your entire viewpoint.

In all these cases, having a user click on a link to a third-party website which is not an ad, the direct commercial value can be difficult to see. Ultimately, what it comes down to is that people implement links when they believe they are referring the user to a high-quality resource on the web and that will bring value to that user. This brings value back to the site implementing the link because the user will have had a good experience on their site because they referred the user to a resource that helped that user, and it may lead to the user to come back for future visits.

The way that Google uses this information is to help it determine which resources are the best quality resources on the web. For example, if someone enters in a query such as "make a rug" Google likely has tens of thousands of pages to choose from that discuss this topic.

How does Google know which one is the best, second best, and so forth? Even the latest AI algorithms can't make that determination by simply analyzing the content. Links help Google see what others on the web think are great resources and act as an input to their algorithms for determining the quality of content.

However, not all links are useful. Ads, of course are biased because they have been paid for. Links that are of low relevance likely count for less, as well as those lacing in any knowledge or expertise about the topic. In addition, there remain many sites that seek to game the link algorithm and drive high rankings for their sites without really deserving them.

In addition to understanding the reasons why some sites might organically implement links to a third party site it is useful to understand what types of

behavior are unnatural and therefore likely to be ignored or penalized by Google.

For example, in the academic world you do not buy placement of a citation in someone else's research paper. Nor do you barter such placements ("I will mention you in my paper if you mention me in yours"), and you certainly would not implement some tactic to inject mentions of your work in someone else's research paper without the writer's knowledge.

You would also not publish dozens or hundreds of poorly written papers just so you could include more mentions of your work in them. Nor would you upload your paper to dozens or hundreds of sites created as repositories for such papers if you knew no one would ever see it there, or if such repositories contained a lot of illegitimate papers that you would not want to be associated with. In principle, you can't vote for yourself.

Of course, all of these examples have happened on the Web with links. All of these practices run counter to the way that search engines want to use links, as they are counting on the links they find being ones that were earned by merit.

This means that search engines don't want you to purchase links for the purpose of influencing their rankings. You can buy ads, of course—there is nothing wrong with that—but search engines would prefer those ad links have the `nofollow` attribute so they know not to count them.

Additionally, pure barter links are valued less or ignored altogether. From 2000 to 2005, it was quite popular to send people emails that offered to link to them if they linked to you, on the premise that this helped with search engine rankings. Of course, these types of links are not real citations either.

Google will not place any value on the links from user-generated content sites, such as social media sites, either. Anywhere people can link to themselves is a place that search engines will simply discount, or even potentially punish if they detect patterns of abusive behavior.

Google invested heavily in developing techniques for detecting poor quality links. For many years it was a highly manual process. However, on April

24, 2012 they took a huge step forward with the initial release of the Penguin algorithm. Penguin was the beginning of their use of automatically detecting links that were poor in quality and either discounting them or assigning an algorithmic penalty on the sites receiving these links.

Penguin used to run separately from the main algorithm and would update only on a periodic basis until the release of Penguin 4.0 on September 23, 2016. As of that date Penguin was fully absorbed into the main algorithm. As of that date Google also changed the algorithm to solely focus on identifying poor quality links and discounting them to zero value. Google's confidence in the efficacy of the Penguin concept had grown high enough that there was no longer a need to penalize these links.

Google's web spam team, however, still manually reviews the link profiles for sites that are considered to have a suspicious link profile and they may assign penalties to those sites. We will discuss this more in the Penalties section of this chapter. For that reason, it is a good idea to have an understanding of what types of links Google doesn't care for.

Links Google Does Not Like

Below is a list of various types of links that Google may consider less valuable, or potentially not valuable at all:

Article directories

> These are sites that allow you to upload an article to them, usually with little or no editorial review. The articles can contain links back to your site, and all you needed to do was upload an article The problem is that this is a form of voting for yourself, and detecting links from these sites is pretty trivial for Google.

Cheap directories

> There are many directories on the Web that exist only to collect fees from as many sites as possible. These types of directories have little or no editorial review, and the owner's only concern is to collect as many listing fees as possible.

Links from countries where you don't do business

If your company does business only in Brazil, there is no reason you should have large numbers of links from Poland and Russia. There is not much you can do if people choose to give you links you did not ask for, but there is certainly no reason for you to proactively engage in activities that would result in your getting links from such countries.

Links from foreign sites with a link in a different language

Some aggressive SEO professionals actively pursue getting links from nearly anywhere. As shown in Figure 7-15, there is no reason to have a "Refinance Your Home Today" link on a page where the rest of the text is in Chinese.

中国可再生能源学会风能专业委员会

（ 中 国 风 能 协 会 ）

CHINESE WIND ENERGY ASSOCIATION

| 首页 | 协会简介 | 行业新闻 | 企业新闻 | 政策法规 |

会员登陆　　　　用户名：　　　　密码：

站内搜索

关键字：

行业新闻

第四届中国可再生能源校融资论坛
关于国家项目~3MW双馈机组
*2009年中国风能新年茶话会
GE风能传动科技与美国第一能源
第二届全国风力发电设备润滑技术
*2008北京国际风能大会暨展
安世亚太"精益研发大会"九城市
南车时代结缘华电新能源，风电项
打造"风电三峡"（张国宝）
2007年中国风电装机世界第三
中德能源与气候政策展望——可再
Canada Expects

最新通知

BTM推出2008世界风电市场报告（WORLD MARKET UP [2009-4-16]
关于举办"海峡两岸风力发电产业交流及合作会议"的通知 [2009-4-13]
关于赴韩国参加"第八届世界风能大会"的通知 [2009-4-2]
关于赴美国参加风能大会及风能考察的通知 [2009-4-1]
关于举办中国新能源与可再生能源装备大会的通知 [2009-3-17]
2009年北京国际风能大会论文征集通知 [2009-3-16]
风能协会招聘信息 [2009-2-20]
关于2009年组团赴国外参加活动的通知 [2009-2-17]
瑞典国际开发署"2009风能开发与利用高级培训班"通知 [2009-2-17]
第八届世界风能大会[WWEC2009]论文征集通知 [2009-2-17]

Refinance your home today!

资料中心 DATUN

Figure 7-15. Foreign language mismatch

Comment spam

Another popular technique in the past was to drop links in comments on forums and blog posts. This practice became much less valuable ever since Google introduced the `nofollow` attribute, but aggressive spammers still pursue it. In fact, they make use of bots that drop comments on an automated basis on blog posts and forums all over the Web. They may post 1 million or more comments this way, and even if only .001% of 1 percent of those links are not `nofollow`ed, it still nets the spammers 1,000 links.

Guest post spam

These are generally poorly written guest posts that add little value for users and have been written just to get a link back to your own site. Consider the example in Figure 7-17, where the author was looking to get a link back to his site with the anchor text "buy cars." He could not even take the time to work that phrase into a single sentence!

St. Mary's

St. Mary's is a town just under an hour's drive from Bicheno, or an hour and a half's drive from Launceston. You can challenge yourself to a climb to the St. Patrick's Head, or the South Sister's Peak, or a relaxing drive in the Elephant Pass - all of which give spectacular views. In town, have a browse through all the craft galleries, bakeries and shops that this small town has to offer.

In terms of getting yourself a vehicle, scope out rental options, or if you're going for an extended holiday, you can consider looking at second hand cars to buy. Cars should be reliable, safe and road worthy of course, so don't buy something that's run down just because it's cheap.

Once you've got a set of wheels, it's time to hit the road!

Figure 7-16. Guest post spam

Guest posts not related to your site

This is a type of guest post spam where the article written does not really relate to your site. If you sell used cars, you should not expect

Google to see any value in a guest post you write about lacrosse equipment that links back to your site. There is no relevance.

In-context guest post links

Another form of guest posting that Google frowns upon is posts that include links in the body of the article back to you, particularly if those links are keyword-rich, and if they don't add a lot of value to the post itself. Figure 8-17 shows a fictional example of what this might look like.

How to Get the Best Travel Deals

- Book early. People who wait until the last minute almost always pay higher prices.

- Use the Internet. Websites such as travel-o-matic can help you save hundreds of dollars on air travel and hotels.

- Try a vacation rental rather than a hotel.

- Look for travel package deals to help cut down the overall cost of both airfare and hotels combined.

This guest post was provided by Sarah Smith an avid traveler who also blogs for www.travel-o-matic.com.

Figure 7-17. Embedded keyword-rich anchor text links

Advertorials

This is a form of guest post that is written like it is an ad. Given the structure, it's highly likely that the site posting it was influenced to do so in some manner. If you are going to include guest posting as part of your strategy, focus on sites that don't permit these types of guest posts.

Guest Posts

While the prior four examples all relate to guest posts, Google more or less frowns on any type of guest posting done for link building. This does not mean you should never guest post, but your goal in doing so should be to get people to read the content you write, and not to get links.

Widgets

One tactic that became quite popular is building useful or interesting tools (widgets) and allowing third-party websites to publish them on their own sites. These normally contained a link back to the widget creator's site. If the content is highly relevant, there is nothing wrong with this idea in principle, but the problem is that the tactic was abused by SEOs, resulting in Google wanting to discount many of these types of links.

Infographics

This is another area that could in theory be acceptable but was greatly abused by SEOs. It is not clear what Google does with these links at this point, but you should create infographics only if they are highly relevant, highly valuable, and (of course) accurate.

Misleading anchor text

This is a more subtle issue. Imagine an example where the anchor text of a link says "information about golf courses," but the page receiving the link is about tennis rackets. This is not a good experience for users and is not something that search engines will like.

Sites with malware

Of course, Google looks to discount these types of links. Sites containing malware are very bad for users, and hence any link from them is of no value, and potentially harmful.

Footer links

Once again, there is nothing inherently wrong with a link from the footer of someone's web page, but as these links are less likely to be clicked on or viewed by users, Google may discount their value. For more on this topic, you can read Bill Slawski's article Google's Reasonable Surfer: How the Value of a Link May Differ Based Upon Link and Document Features and User Data" (http://bit.ly/reasonable_surfer), which discusses Google's "reasonable surfer" patent.

Links in a list with unrelated links

This can be a sign of a purchased link. Imagine you find a link to your "Travel Australia" website mixed in a list of links with an online casino site, a mortgage lead generation site, and a lottery ticket site. This does not look good to Google.
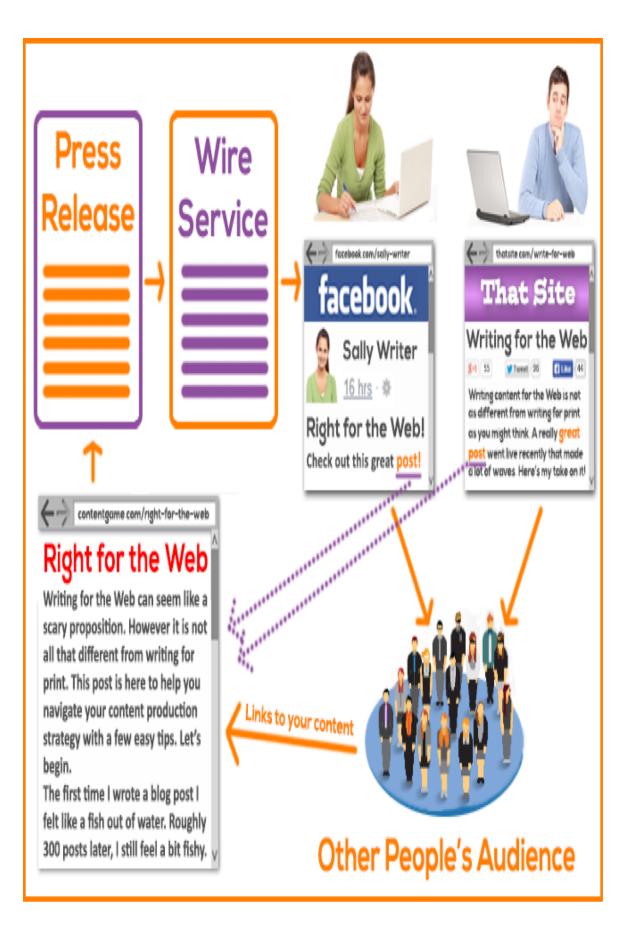
Links from poor-quality sites

The links that have the most value are the ones that come from very high-quality sites that show substantial evidence of strong editorial control. Conversely, as quality drops, editorial control tends to as well, and Google may not count these links at all.

Press releases

It used to be quite popular to put out lots of press releases, complete with keyword-rich text links back to your site. Of course, this is a form of voting for yourself, and this is not the way that press releases should be used to promote your site. As shown in Figure 7-18, a much better

way to use press releases is to get your news in front of media people and bloggers, and hope that it's interesting enough that they will write about you or share your news on social media.

Press Release

Wire Service

facebook

facebook.com/sally-writer

Sally Writer

16 hrs · ☼

Right for the Web!

Check out this great post!

That Site

thatsite.com/write-for-web

Writing for the Web

g+1 55   Tweet 96   f Like 44

Writing content for the Web is not as different from writing for print as you might think. A really great post went live recently that made a lot of waves. Here's my take on it!

---

contentgame.com/right-for-the-web

Right for the Web

Writing for the Web can seem like a scary proposition. However it is not all that different from writing for print. This post is here to help you navigate your content production strategy with a few easy tips. Let's begin.

The first time I wrote a blog post I felt like a fish out of water. Roughly 300 posts later, I still feel a bit fishy.

Links to your content

Other People's Audience

Figure 7-18. The right way to use press releases

Bookmark sites

> There are many quality sites for saving interesting links for your own benefit, such as Delicious, Evernote, Diigo, and many others. However, as these are user-generated content sites, their links are `nofollow`ed and thus have no value in ranking your site.

Not all of the types of links in the preceding list will necessarily result in your site being penalized, but they are all examples of links that Google will likely not want to count.

Cleaning Up Low Quality Backlinks
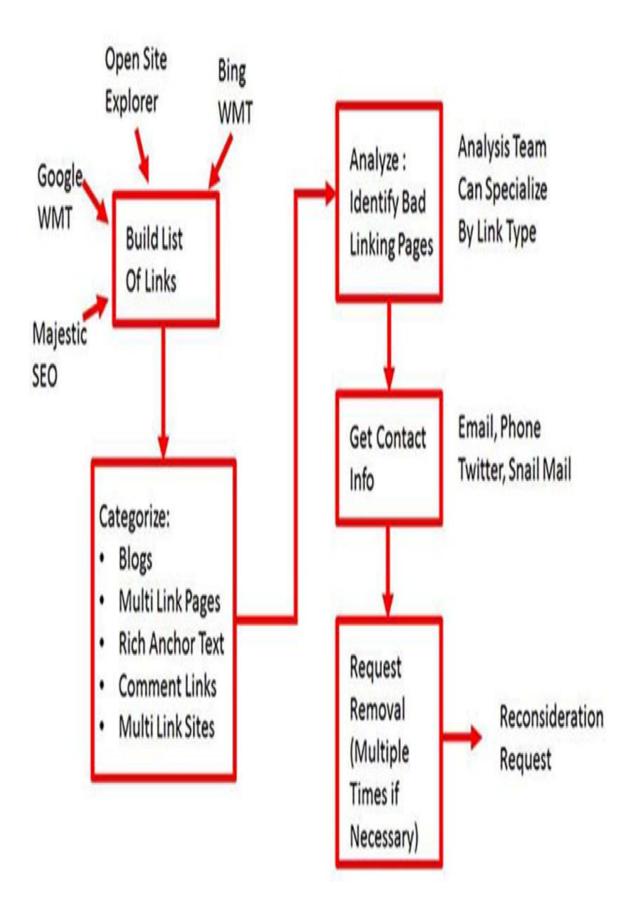
The first part of the link cleanup process is to establish the right mindset. As you review your backlink profile, consider how Google looks at your links. Here are some rules of thumb to help you determine whether a link has real value:

- Would you want that link if Google and Bing did not exist?

- Would you proudly show it to a prospective customer right before she is ready to buy?

- Was the link given to you as a genuine endorsement?

As you review your backlinks, you may find yourself at times trying to justify a link's use. This is usually a good indicator that it's *not* a good link. High-quality links require no justification—it's obvious that they are good links.

Another key part of this process is recognizing the need to be comprehensive. Losing a lot of your traffic is scary and being impatient is natural. If there is a manual link penalty on your site you will be anxious to send in your reconsideration request, but as soon as you do, there's nothing you can do but wait.

If you don't do enough to remove bad links, Google will reject your reconsideration request, and you have to go through the whole process again. If you end up filing a few reconsideration requests without being successful, Google may send you a message telling you to pause for a while. Make a point of being very aggressive in removing and disavowing links, and don't try to save a lot of marginal ones. This almost always speeds up the process in the end. In addition, those somewhat questionable links that you're trying to save often are not helping you much anyway.

With all this in mind, you also want to be able to get through the process as quickly as possible. Figure 7-19 is a visual outline of the link removal process.

Figure 7-19. Process for removing links

Precategorizing the links is quite helpful in speeding up this process. For example, you can identify many of the blogs simply by using the Excel filter function and filtering on "blog." This will allow you to more rapidly review the links for problems. Tools such as Remove'em (http://www.removeem.com/) and Link Detox (http://www.linkdetox.com/) will do this for you as well.

This step is especially helpful if you know you have been running an aggressive guest posting campaign, or worse, paying for guest post placements. Some additional tips include:

- You do not need to worry about links that are marked as `nofollow`.

- Links from sites with very low PageRank for their home page probably are adding little to no value to your site.

- Links from very low-relevance pages are not likely to be adding much value either.

In addition, contacting sites directly and requesting that they remove links can be quite helpful. Google likes to see that you're putting in the effort to clean up those bad links.

Remember that reconsideration requests are reviewed by members of the webspam team at Google. This introduces a human element that you can't ignore. The members of this team make their living dealing with sites that have violated Google's guidelines, and you are one of them.

As we note in the following section, "Sources of Data for Link Cleanup", even when you use all the available sources of link data, the information you have is incomplete. This means that it's likely that you will not have removed all the bad links when you file your reconsideration request, even if you are very aggressive in your review process, simply because you don't have all the data. Showing reviewers the good faith effort to remove some

of the bad links is very helpful, and can impact their evaluation of the process.

However, there is no need to send link removal requests to everyone in sight. For example, don't send them to people where the link to you is marked with `nofollow`.

Once the process is complete, if you have received a manual penalty, you are ready to file a reconsideration request.

Sources of Data for Link Cleanup

Google provides a list of external links in the Search Console account for your site. Figure 7-20 shows a sample of that report.

Figure 7-20. Search Console links report

The problem with this list is that it tends to be incomplete, thus we recommend that you also pull links from several other sources. Some of the best additional sources include Ahrefs (https://ahrefs.com/), Majestic SEO (https://www.majestic.com), SEMrush (https://www.semrush.com), Link Explorer (https://moz.com/link-explorer), and LinkResearchTools (https://www.linkresearchtools.com).

As with Search Console, each of these tools only offers a partial list of the link. In the case of these software vendors they are relatively small and the challenge of crawling the web as thoroughly as Google is a large one so it should be no surprise that they don't cover the entire web.

However, building a database of the combination of data from all of these tools will show a more complete list of links. During the course of performing a study on the link tool vendors Perficient found that using these data sources together resulted in finding twice as many links when compared to the one vendor that reported the most links (https://blogs.perficient.com/2021/01/26/study-who-has-the-largest-index-of-links/).

Of course, there will also be a lot of overlap in what they show, so make sure to deduplicate the list. However, even the combination of all these sources is not comprehensive. Google shares only a portion of the links it is aware of in Search Console. The other link sources are reliant on the crawls of their individual companies, and crawling the entire Web is a big job that they simply do not have the resources for.

Using Tools for Link Cleanup

There are tools available to help speed up link removal by automating the process of identifying bad links. These include Remove'em (https://www.removeem.com/) and Link Detox (https://smart.linkresearchtools.com/new/link-detox). These tools can potentially help you identify some of your bad links. However, it is a good idea to not rely solely on these tools to do the job for you.

Each tool has its own algorithms for identifying problem links, and this can save you time in doing a full evaluation of all your links. However, keep in mind that Google has spent more than 15 years developing its algorithms for evaluating links and it is a core part of its business to evaluate them effectively, including detecting link spam. Third-party tools won't be as sophisticated as Google's algorithm. They can detect some of the bad links, but not necessarily all of the ones you will need to address. You should plan on evaluating all of the links—not only the sites labeled as toxic, but also any that are merely suspicious or even innocuous. Use your own judgment, and don't just rely on the tools to decide for you what is good or bad.

The Disavow Links tool

Google provides a tool to allow you to disavow links (http://bit.ly/disavow_links). The Disavow Links tool tells Google that you no longer wish to receive any PageRank (or other benefit) from certain links. This gives you a method for eliminating the negative impact of bad links pointing to your site. Figure 7-21 shows what the tool's opening screen looks like.

Google Search Console

Disavow links to your site

If you have a manual action against your site for unnatural links, or if you think that you're about to get one because of paid links or link schemes that violate our quality guidelines, ask the other site to remove those links. If you can't get these links removed, then disavow those sites using this tool. More information

⚠ This is an advanced feature and should only be used with caution ⌄

Select a property to view its status or to upload a new list of disavowed links

▼

Domain properties are not supported at this time.

Figure 7-21. Opening screen of Google's Disavow Links tool

Note that Google includes the following text: "You should still make every effort to clean up unnatural links pointing to your site. Simply disavowing them isn't enough." This is good advice, as Google employees who review reconsideration requests like to see that you have invested time in getting the bad links to your site removed.

Once you select a site (blacked out in Figure 7-21) and click the Disavow Links button, you are taken to another screen that includes the following warning:

This is an advanced feature and should only be used with caution. If used incorrectly, this feature can potentially harm your site's performance in Google's search results. We recommend that you only disavow backlinks if you believe that there are a considerable number of spammy, artificial, or low-quality links pointing to your site, and if you are confident that the links are causing issues for you.

On this screen, you need to click Disavow Links again, after which you'll be taken to a third and final screen, shown in Figure 7-22.

Disavow Links

This is an advanced feature and should only be used with caution. If used incorrectly, this feature can potentially harm your site's performance in Google's search results. We recommend that you only disavow backlinks if you believe that there are a considerable number of spammy, artificial, or low-quality links pointing to your site, and if you are confident that the links are causing issues for you.

The file below contains the list of disavowed links. To edit the list, click **Download**.

████████.txt   Download   Delete

Results for the submission on April 3, 2014 11:34:20 AM UTC-4

You successfully uploaded a disavow links file (████████.txt) containing 75 domains and 9065 URLs.

Choose File

Submit   Done

Figure 7-22. Final screen of Google's Disavow Links tool

The sample screen in Figure 7-22 shows that the current disavow file for this site (the name is blacked out) is disavowing 75 domains and 9,065 URLs.

Considering that the link data you have is incomplete (as described in "Sources of Data for Link Cleanup"), it is best practice to disavow entire domains. In other words, if you see one bad link coming to you from a domain, it is certainly possible that there are other bad links coming to you from that domain, and that some of these bad links are not in the data available to you.
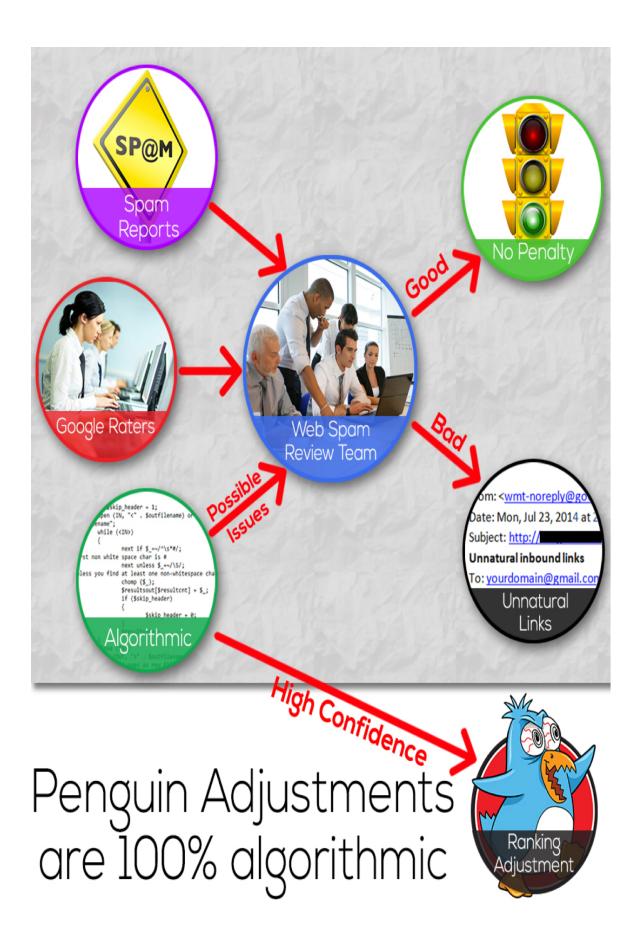
An example would be a guest post that you want to disavow. Perhaps you have done only one guest post on that site, but the post will also appear in category pages and date-based archives on that blog. If you disavow only the post page, you still have many other bad links from that site that Google has found.

In the example shown in Figure 7-22, it is quite likely that this publisher has not solved his problem and that many (if not all) of the disavowed URLs should be made into full disavowed domains. That is usually the safest course of action.

Refer to the Google help page for more specifics on formatting the disavow file (http://bit.ly/disavow_backlinks).

Google Manual Actions (Penalties)

There are two ways that you can lose traffic, algorithm changes by Google and manual actions. Algorithm changes are not penalties and do not involve any human component, whereas manual penalties do. While the details of what prompts Google to perform a manual review of a website are not always evident, there appear to be several ways that manual reviews can be triggered. Figure 7-23 illustrates how a manual review might be triggered in the case of a site that has problems with its link profile.

Spam Reports

No Penalty

Good

Google Raters

Web Spam Review Team

Bad

Possible Issues

```
skip_header = 1;
open (IN, "<" . $outfilename) or
ename";
    while (<IN>)
    {
        next if $_=~/^\s*#/;
st non white space char is #
        next unless $_=~/\S/;
less you find at least one non-whitespace cha
        chomp ($_);
        $resultsout[$resultcnt] = $_;
        if ($skip_header)
        {
            $skip_header = 0;
```

Algorithmic

om: <wmt-noreply@go

Date: Mon, Jul 23, 2014 at 2

Subject: http://

Unnatural inbound links

To: yourdomain@gmail.con

Unnatural Links

High Confidence

Penguin Adjustments are 100% algorithmic

Ranking Adjustment

Figure 7-23. Possible ways that Google manual reviews may be triggered

Note that in some cases an algorithm may trigger an algorithmic ranking adjustment (as shown in Figure 7-23, algorithmic adjustments are made only when Google's confidence in the signals is very high; if the confidence level is not high but there are indications of a problem, a manual review might be initiated), however these are not considered "penalties" by Google . Here is a summary of the major potential triggers:

Spam report

Any user (including your competitor) can file a spam report in Google (http://bit.ly/report_webspam). While Google has not revealed how many of these it receives on a daily basis, it's likely that they receive large volumes of these reports. Google evaluates each report, and if it finds one credible (it may run some type of algorithmic verifier to determine that), then it conducts a manual review.

Algorithmically triggered review

While this approach has never been verified by Google, it's likely that Google uses algorithms to trigger a manual review of a website. The premise is that Google uses algorithms that identify large quantities of sites whose behavior is potentially bad, but not bad enough for Google to algorithmically penalize them, so these sites would be queued for manual review. Google could also implement custom algorithms designed to flag sites for review.

Regular search results reviews

Google maintains a large team of people who perform manual reviews of search results to evaluate their quality. This effort is primarily intended to provide input to the search quality team at Google that they can use to help them improve their algorithms. However, it is quite possible that this process could also be used to identify individual sites for further scrutiny.

Once a review is triggered, the human reviewer uses a set of criteria to determine if a penalty is merited. Whatever the outcome of that review, it is likely that Google keeps the notes from the review in a database for later use. Google most likely keeps a rap sheet on all webmasters and their previous infractions, whether they result in a penalty or not.

GOOGLE SEARCH CONSOLE MESSAGES

Since April 2012, Google has maintained a policy of sending all publishers that receive a manual penalty a message in their Search Console describing the nature of the penalty. These messages describe the penalty in general terms, and it is up to the publisher to figure out how to resolve it. Generally, the only resource that Google provides to help with this is its Webmaster Guidelines (https://bit.ly/webmaster_best_practices). If you receive such a message, then the reconsideration request option in Google Search Console becomes available.

Types of Manual Actions/Penalties

Manual penalties come in many forms. The most well-known types of penalties are thin content or link related, but you can also get a variety of other penalties. Some of the most common types of manual penalties are discussed in the following sections. Google provides two key pages to help you understand the different types of penalties and what they mean:

1. Manual Actions Report: https://support.google.com/webmasters/answer/9044175?hl=en.

2. Security Issues Report: https://support.google.com/webmasters/answer/9044101?hl=en

Familiarity of the content of these two pages is an important aspect of your SEO program as they detail the types of behaviors that cause Google to

have concerns with your site. Here are some of the more common types of issues that sites run into:

Thin-content penalties

This penalty relates to pages that don't add enough value to users in Google's opinion. Figure 7-24 shows an example of a thin-content message from Google in Search Console.

Figure 7-24. Thin-content penalty message

Unfortunately, when you receive this type of penalty, Google doesn't provide any guidance on what the cause might be. It does tell you that it is a thin-content penalty, but the rest is up to you. There are four primary triggers for thin-content penalties:

Pages with little useful content

As the name of the penalty suggests, pages with very little content are potential triggers for this penalty. This is especially true if there are a large number of these pages, or if there is a particular section on the site that has a significant percentage of its pages deemed thin.

Thin slicing

This happens to publishers who implement pages that are really designed to just garner search traffic. What these publishers often do is build pages for each potential search query a visitor might use, even if the variations in the content are quite small or insignificant. To use an

earlier example, imagine a site with information on nursing schools with different pages with the following titles:

- Nursing schools

- Nursing school

- Nursing colleges

- Nursing universities

- Best nursing schools

Sometimes publishers do this unintentionally, by autogenerating content pages based on queries people enter when using the search function for the website. If you decide to do something like this, then it's critical to have a detailed review process for screening out these thin-slicing variants, pick one version of the page, and focus on it.

Doorway pages

These are pages that appear to be generated just for monetizing users arriving from search engines. One way to recognize these types of pages is that they are usually pretty much standalone pages with little follow-on information available, and/or they are pages that are largely written for search engines and not users.

The user arriving on these pages basically has two choices: buy now or leave.

Poor integration into the overall site

Another issue to look for is whether parts of your site are not well integrated into the rest of the site. Is there a simple way for users to get to these pages from the home page, from the main navigation of the site, or at least from a major section of the site?

If you have a section that appears to be isolated from the rest of your site, that could result in a thin-content penalty.

Once you believe you have resolved these issues, you need to submit a reconsideration request. You can read more about this in the section below, "Filing Reconsideration Requests". Once you have filed this request, you simply wait until Google provides a response. This process normally takes two to three weeks.
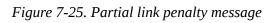
If you are successful, then you are in good shape and just need to make sure not to overstep your boundaries again in the future. Otherwise, it's back to the drawing board to see what you might have missed.

Partial link penalties

Another possible manual penalty is a partial link penalty. This is sometimes called an "impacts links" penalty, as that term is part of the message you get from Google (see Figure 7-25). These penalties indicate that one or a small number of your pages have been flagged for bad linking behavior.

Normally, only the rankings and traffic for those particular pages suffer as a consequence of this penalty.

Partial matches

▼ Some manual actions apply to specific pages, sections, or links

Reason	Affects
Unnatural links to your site—impacts links Google has detected a pattern of unnatural artificial, deceptive, or manipulative links pointing to pages on this site. Some links may be outside of the webmaster's control, so for this incident we are taking targeted action on the unnatural links instead of on the site's ranking as a whole. Learn more.	Some incoming links

REQUEST A REVIEW

Figure 7-25. Partial link penalty message

Unfortunately, Google does not tell you which of your pages is receiving the penalty, so you have to determine that for yourself. This penalty is normally caused by too many questionable or bad links to pages other than your home page.

The cause is often a link-building campaign focused on bringing up the rankings and search traffic to specific money pages on your site. One of the more common problems is too many links with keyword-rich anchor text pointing to those pages, but other types of bad links can be involved as well. The steps to recover from this type of penalty are:

1. Pull together a complete set of your links as described in the section "Sources of Data for Link Cleanup".

2. Look for pages on your site, other than the home page, that have the most links.

3. Examine these pages for bad links as described in the section "Links Google Does Not Like".

4. Use the process described in the section "Bad Link Cleanup Process".

5. Submit a reconsideration request as described in the section "Filing Reconsideration Requests".

Once you have sent in the reconsideration request, the only thing you can do is wait. As noted previously, it normally takes two to three weeks before you get a response. Google will either let you know you have succeeded and confirm that it has removed the penalty, or it will tell you that you have failed, in which case you have to take a deeper look at your links and figure out what you missed in your previous attempt.

Sitewide link penalties

Manual link penalties can also be applied on a sitewide basis. This usually means more than a few pages are involved, and may well also involve the home page of the site. With this type of penalty, rankings are lowered for the publisher on a sitewide basis.

As a consequence, the amount of lost traffic is normally far more than it is for a partial link penalty. Figure 7-26 shows an example of a sitewide link penalty message.

☆ ⚠ http:/ ██████████ Unnatural inbound links

Google has detected a pattern of artificial or unnatural links pointing to your site. Buying links or participating in link schemes in order to manipulate PageRank are violations of Google's Webmaster Guidelines.

As a result, Google has applied a manual spam action to ██████████ There may be other actions on your site or parts of your site.

Recommended action

- Use the Links to Your Site feature in Webmaster Tools to download a list of links to your site.

- Ensure that unnatural links pointing to your site are removed.

- When these changes are made, and you are satisfied that links to your site follow Google's Webmaster Guidelines, submit a reconsideration request. If you're unable to remove links pointing to your site, please provide as much detail as possible in your reconsideration request.

- For an updated list of manual actions currently applied to your site, visit the Manual Actions page. If no manual actions are listed, there is no longer a need to file a reconsideration request.

If we determine your site is no longer in violation of our guidelines, we'll revoke the manual action.

If you have any questions about how to resolve this issue, please visit the Webmaster Help Forum.

Figure 7-26. Sitewide link penalty message

The steps to recover from this type of penalty are the same as previously outlined in the Partial Link Penalties section of this chapter.

Once you have sent in the reconsideration request, the only thing you can do is wait. It normally takes two to three weeks before you get a response. Google will either let you know you have succeeded and confirm that it has removed the penalty, or it will tell you that you have failed, in which case you have to take a deeper look at your links and figure out what you missed in your previous attempt.

Other Types of Manual Actions/Penalties

Some of the other manual penalties include:

Cloaking and/or sneaky redirects

> You can get this message if Google believes you are showing different versions of pages to Googlebot than you show to users. To diagnose this, use the URL Inspector tool in Search Console to retrieve the page. Use the tool to load the same page in another browser window and compare the two pages. If you don't have access to Search Console the next best bet is the Mobile Friendly Test Tool ().
>
> If you see differences, invest the time and effort to figure out how to remove the differing content. You should also check for URLs that redirect and send people to pages that are not in line with what they expected to see—for example, if they click on anchor text to read an article about a topic of interest but instead find themselves on a spammy page trying to sell them something.
>
> Another potential source of this problem is *conditional redirects*, where users coming from Google search, or a specific range of IP addresses, are redirected to different pages than other users.

Hidden text and/or keyword stuffing

This message is generated if Google believes you are stuffing keywords into your pages for the purpose of manipulating search results—for example, if you put content on a page with a white background using a white font, so it is invisible to users but search engines can still see it.

Another way to generate this message is to simply repeat your main keyword for a page over and over again in hopes of influencing search results.

User-generated spam

This type of penalty is applied to sites allowing user-generated content (UGC) that are perceived to not be doing a good job of quality control on that content. It's very common that sites with UGC become targets for spammers uploading low-quality content with links back to their own sites.

The short-term fix for this is to identify and remove the spammy pages. The longer-term fix is to implement a process for reviewing and screening out spammy content to prevent it from getting on your site in the first place.

Unnatural links from your site

This is an indication that Google believes you are selling links to third parties, or participating in link schemes, for the purposes of passing PageRank. The fix is simple: remove the links on your site that look like paid links, or add a `nofollow` attribute to those links.

Master Security Issues Report

Google will communicate this penalty by sending you a message in Search Console and/or by showing indications that your site has been hacked (and is dangerous to visit) in the search results. The most common cause for this penalty is failing to keep up with updates to your content management system (CMS).

Spammers take advantage of vulnerabilities in the CMS to modify your web pages, most often for the purpose of inserting links to their own sites, but sometimes for more nefarious purposes such as accessing credit card data or other personally identifiable information.

To resolve the problem, you will need to determine how your site has been hacked. If you don't have technical staff working for you, you may need to get help to detect and repair the problem. To minimize your exposure going forward, always keep your CMS on the very latest version possible.

Pure spam

Google will give you this message in Search Console if it believes that your site is using very aggressive spam techniques. This can include things such as automatically generated gibberish or other tactics that appear to have little to do with trying to add value for users.

If you get this message, there is a strong chance that you should simply shut down the site and start with a new one.

Spammy freehosts

Even if your site is clean as a whistle, if a large percentage of the sites using your hosting company are spamming, Google may take action against all of the sites hosted there. Take care to make sure you are working with a highly reputable hosting company!

For any of these problems, you need to address the source of the complaints. When you believe you have done so, follow the procedure outlined in the section "Filing Reconsideration Requests".

Diagnosing the Cause of a Traffic Loss

The first step to identifying the cause of a traffic loss is to check your analytics data to see if the drop is in fact a loss of organic search engine

traffic. If you have Google Analytics, Adobe Analytics, or other analytics package on your site, make sure you check your traffic sources, and then isolate just the Google traffic to see if that is what has dropped.

If you confirm that it is a drop in Google organic search traffic, then the next step is to check if you have received a message in Google Search Console indicating that you have been penalized by Google. Figure 7-27 shows an example of a manual penalty message from Google. This is what Google refers to as a *manual action*.

View messages for http:/█████████ »

⤺   Delete

☆ Reconsideration request for http:/██████████ : Site violates Google's quality guidelines

Dear site owner or webmaster of http:/██████████

We received a request from a site owner to reconsider http:/██████████ for compliance with Google's Webmaster Guidelines.

We've reviewed your site and we still see links to your site that violate our quality guidelines.

Specifically, look for possibly artificial or unnatural links pointing to your site that could be intended to manipulate PageRank. Examples of unnatural linking could include buying links to pass PageRank or participating in link schemes.

We encourage you to make changes to comply with our quality guidelines. Once you've made these changes, please submit your site for reconsideration in Google's search results.

If you find unnatural links to your site that you are unable to control or remove, please provide the details in your reconsideration request.

If you have additional questions about how to resolve this issue, please see our Webmaster Help Forum for support.

Sincerely,

Google Search Quality Team

Figure 7-27. Manual Action/Penalty message in Google Search Console

If you have received such a message, you now know what the problem is and you can get to work fixing it. It is not fun to have the problem but knowing what you are dealing with is the first step in recovery.

If you don't have such a message, you will need to dig deeper to determine the source of your problem. The next step is to determine the exact date on which your traffic dropped. There are many tools on the market that you can then use to see if there were significant Google updates on that day. Here are eight possible tools that you can use for this purpose:

- Mozcast - https://moz.com/mozcast/

- Moz Google Algorithm Update History - https://moz.com/google-algorithm-change

- SEMrush Sensor - https://www.semrush.com/sensor/

- RankRanger Rank Risk Index Tool - https://www.rankranger.com/rank-risk-index

- Accuranker 'Grump' Rating - https://www.accuranker.com/grump

- Algoroo - https://algoroo.com/

- Advanced Web Rankings Google Algorithm Changes - https://www.advancedwebranking.com/google-algorithm-changes/

- Cognitive SEO Signals - https://cognitiveseo.com/signals/

For example, if your site traffic dropped on February 10, 2021, Figure 7-28 suggests that sites that suffered traffic losses on this date may have been impacted by Google's Passage Indexing Update.

Passage Indexing (US/English) — February 10, 2021   Confirmed

Google rolled out so-called "passage indexing" (which is probably closer to passage ranking) for US/English queries. While we measured two days of moderate

rankings flux, it was unclear exactly how the update impacted SERPs. Google initially estimated this update would impact 7% of queries.

- Google passage ranking now live in US English search results (SEL)
- How AI is powering a more helpful Google (Google)

Figure 7-28. Moz Google algorithm update page showing the Passage Indexing Update

If you haven't gotten a message in Google Search Console, and the date of your traffic loss does not line up with a known Google algorithm update, the process of figuring out how to recover is much harder, as you don't know the reason for the drop.

Google does make smaller changes to its algorithms on a daily basis. From its perspective, these are smaller changes and not major updates. However, even these could possibly have a significant impact on traffic to your site, either positive or negative.

If they do impact you negatively, such tweaks may be much harder to recover from as well. The best strategy is to focus on the best practices outlined in Chapter 6 and Chapter 7, or if you can afford SEO advice, bring in an expert to help you figure out what to do next.

Part of the reason Google makes daily tweaks is that it allows them to make small improvements on a continuous basis as well as run a variety of tests for purposes of improving the algorithm. Sometimes the scope of these changes rises to a level that the industry recognizes them and you can see active discussions about what's taking place on Twitter or in the major search industry journals such as Search Engine Land, Moz, Search Engine Journal and others.

Some of these updates get confirmed by Google and others do not. Nonetheless, any of them can impact traffic to your site in material ways.

Filing Reconsideration Requests to Remediate Manual Actions/Penalties

Reconsideration requests are applicable only to penalties. Unless you have a manual penalty you will not be able to file one to address traffic losses you may have experienced. The next thing to realize about your reconsideration request is that a person will review it, and that person likely reviews large numbers of them every single day. Complaining about what has happened to your business, or getting aggressive with the reviewer, is not going to help your cause at all.

The best path is to be short and to the point:

1. Briefly define the nature of the problem. Include some statistics if possible.

2. Explain what went wrong. For example, if you were ignorant of the rules, just say so, and tell them that you now understand. Or, if you had a rogue SEO firm do bad work for you, say that.

3. Explain what you did to clean it up:

   - If you had a link penalty, let them know how many links you were able to get removed.

   - If you did something extraordinary, such as removing and/or disavowing all of your links from the past year, tell

them that. Statement actions such as this can have a strong impact and improve your chances of success.

4. 4. Clearly state that you intend to abide by the Webmaster Guidelines going forward.

As already noted, keep your reconsideration request short. Briefly cover the main points and then submit it using the Search Console account associated with the site that received the penalty. In fact, you can't send it from an account without a manual penalty.

Expected reconsideration request timeline

Once you have filed the request, you now get to wait. The good news is that you generally get a response in two to three weeks. Hopefully, you will be successful! If not, you have to go back to the beginning of the process to figure out what you missed.

Recovering from Traffic Losses Not Due to a Manual Action/Penalty

Reconsideration requests are only applicable in the case that you have been subject to a penalty. For all other causes of lost traffic all you can really do is to make the improvements to your site that you believe will help you recover and wait.

Google has to recrawl your site to see what changes you have made. If you have made sufficient changes it may still take Google several months before it has seen enough of the changed or deleted pages to tilt the balance in your favor.

What if you don't recover?

Sadly, if your results don't change, this usually means that you have not done enough to address whatever issues caused your traffic loss. Don't overlook the possibility that your development team may have made changes that cause your site to be difficult for Google to crawl. Perhaps they made a change to the platform the site is implemented on, used

JavaScript in a way that hides content from Google, blocked content from being crawled in Robots.txt or some other technical issue.

If this is not the nature of the issue, then you will need to keep investing in the areas of your site that you might think are related to the traffic drop, or more broadly, that will help increase the value of your site. Address this situation by taking on the mission to make your site one of the best on the Web.

This requires substantial vision and creativity. Frankly, it's not something that everybody can accomplish without making significant investments of time and money. One thing is clear: you can't afford to cut corners when trying to address the impact of traffic losses from Google.

If you continue to invest a lot of time and made many improvements, but you still have content that you know is not so great, or other aspects of the site that need improvement, chances are pretty good that you haven't done enough. You may find yourself four months later wishing that you had kept at the recovery process.

In addition, the Google algorithm is constantly evolving. Even if you have not been hit by traffic loss, the message from Google is clear: it is going to give the greatest rewards to sites that provide fantastic content and great user experiences. Thus, your best path forward is to be passionate about creating a site that offers both. This is how you maximize your chances of recovering from any traffic loss, and from being impacted by future Google updates.

Conclusion

Traffic losses due to manual actions/penalties or algorithmic updates can have a significant impact on your business. It is therefore critical as a digital marketer to understand Google's ever-evolving Webmaster Guidelines (http://bit.ly/webmaster_best_practices), to create compelling websites that satisfy the needs of the end user, and to promote these websites with legitimacy and longevity in mind.

CHAPTER EIGHT

Auditing and Troubleshooting

Even if you have a mature SEO organization, new challenges and/or opportunities can still arise. The reasons for this are many, including:

- The technology you use to implement your site requires you to work around its limitations to properly support SEO.

- Many (or most) of the people in your organization don't understand how SEO works, or worse, don't value SEO, leading to mistakes to be made.

- Google algorithm changes can create new opportunities or issues to address.

- Competition can invest heavily in SEO resulting in new challenges to your SEO market share.

As a result, knowing how to conduct an SEO audit and being able to troubleshoot SEO problems are essential skills for any professional SEO.

SEO Auditing

There are many reasons why you may need to conduct an SEO audit. Perhaps you plan to perform one every quarter, or every six months. Perhaps a Google algorithm update impacted your site traffic. Or maybe you saw a drop in your organic search traffic in your analytics data. It could also be that you are proactively trying to find ways to better optimize your site in order to increase organic traffic market share.

Regardless of how you think of why you're doing the audit, the underlying purpose of performing it comes down to finding ways to improve the SEO of your site and therefore increase the organic search traffic to the site. It's good to keep that in mind throughout the entire process and to remind all impacted stakeholders as well.

For this reason this chapter focuses on finding SEO issues and opportunities, either as a result of technical SEO problems on your site, through the creation of new content or making improvements to existing content (You can read more about the broader goal of creating an SEO strategy in Chapter 4). Once you've built a list of problems and opportunities the next step is to devise a plan to make changes to improve overall organic search results.

Because of the dynamic nature of the environment that we live in it's a good idea to conduct audits on a regular basis. For example, you might decide to schedule an audit once per quarter or twice per year. This type of cadence will enable you to limit the amount of tech debt driven SEO problems that your site accumulates. The regular audit schedule will offer other benefits as well, which is that it can help raise the consciousness of your entire organization around the impact and importance of SEO.

Unscheduled Audits

Having regularly scheduled audits is smart business but the need can also arise from time to time for an unscheduled audit. Some reasons why this may occur include:

- A Google algorithm update

- An unexpected organic search traffic drop

- An unexpected rankings drop

- Site changes being pushed live without being checked for SEO first

- A Senior manager demands some type of review

- A merger or some other corporate activity causes a need to merge web properties together

- A site redesign takes place

- New brand guidelines necessitate significant site changes

If any of these events occur you may feel the need to do an immediate audit. In such a case the focus of your audit will be guided by the event causing you to do it. For example, if you have experienced an unexpected rankings drop for one page or across a single section of a site. In such an event your audit will likely focus on aspects of your site that are most likely to impact that ranking.

Customized Approaches to Audits

There are many types of variations of SEO audits that you can conduct. Sometimes this could be because the audit was unscheduled and therefore focused on a specific issue. Some of the reasons why you may conduct a subset of a full audit include:

- Routine analytics monitoring may show specific issues you want to investigate. For example, perhaps traffic to one specific URL may have dropped and your audit may focus on issues specific to that page.

- You may have a new version of the site that you are about to release and are checking to ensure that new problems have not been introduced. In this event it is very useful to perform a technical SEO check only.

- A new Google algorithm may have gotten released and industry commentary on the release may expose specific issues that the new algo may focus on. In such a case, you should consider a mini-audit focusing on whether or not your site has those issues.

- You discover that traffic to different language and country versions of your site are lagging far behind. If this happens you may focus an audit just on the international pages and how they are integrated into your overall web ecosystem. This should include verifying that your hreflang tags are correctly implemented across all different language and country versions of the site (or sites if your language and country versions appear on different domains).

- The web site has Page Experience issues. This Google algorithm (which is discussed more in Chapter 6) includes the Core Web Vitals (Cumulative Layout Shift, First Input Delay, and Largest Contentful Paint) as well as whether or not the site is mobile friendly, uses interstitials, and serves pages over HTTPS. Audits of these are easily done separately as they often involve different teams to do the work.

- Changes are being proposed to the main navigation of the site. This can have a large impact on SEO and there are many reasons beyond SEO why you may want to change this, including usability and user experience concerns.

- Brand guidelines have been updated. This would result in needing a review of how your brand is represented across the site. In this type of audit you will want to validate that SEO for the site is not being damaged in the process.

- Marketing may want to have a brand audit of the site done. During this process you would focus on validating your brand guidelines and in the process make sure that any corrections made don't harm SEO.

- The organization acquires another business or web site. You may want to combine those new web assets into an existing site or establish cross-linking between the web properties. These types of activities should be optimized for optimal SEO.

- Multiple departments with responsibility for different portions of the site. This can lead to frustrating outcomes if different groups at your organization make changes to the site without first validating that SEO has been actively considered in the process. You may find out a change has been pushed live without SEO being considered and you need to retroactively check it out.

- The site has a paywall and it's correctly implemented from a user perspective but it is not performing in organic search how you want it to. It may be that your implementation of flexible sampling isn't implemented properly. You can read more about flexible sampling in Chapter 6.

- Your organization may be making changes to the site to improve accessibility of the site to those with some form of disability, such as being blind, reduced mobility, and cognition (in the United States these issues are addressed by a civil rights law known as The Americans with Disabilities Act of 1990 – other countries have different standards). If needed, improving site accessibility is important for your business, but there can be significant interactions with SEO optimization of the site.

- Some organizations set up a rotation of regularly scheduled partial audits for the web site. For larger web properties this is an excellent idea to do and it can allow you to balance the workload of having to conduct a full audit of everything all at once.

- Due to resource limitations you want to focus on an audit that is tackling your biggest areas of concern.

These are just some examples of reasons you may want to perform different types of audits and there are many other reasons why you may want to focus on auditing specific issues. Once the need to validate some portion of the site via a partial audit the key is to determine what factors are of most importance based on the actual need your trying to address.

Pre-Audit Preparations

Prior to beginning an audit you should take several steps to lay the groundwork for a successful result. This preparation will help ensure that the audit focuses on the right areas and that the work goes as smoothly as possible. Key steps at this stage are:

- Identify the reasons why you are deciding to conduct an audit. Is it a routine check that you perform quarterly or twice per year? Or is it driven by some event such as a drop in organic search traffic? Are there specific issues that you are looking for? Make a clear list of the items you're looking for and use that to guide the specific areas where your audit focuses.

- Connect with the various stakeholders that are impacted by the audit and understand what they might be looking to learn from the audit or what their concerns might be. Be sure to include any development teams or content teams who might expect to receive work requests to support the audit.

- Define the scope of the audit. Are you crawling the entire web site or just a portion? Is an SEO content audit part of the plans? Are you covering the entire site or just a given section?

- Develop an SEO audit plan that includes what steps will be taken and in what order.

- Review the proposed plan with the stakeholders and get their buy-in to the plan.

- Identify the people who will lead and conduct the audit and confirm their availability to do the work.

- Nail down which special tools or databases that may be required to support the audit and ensure that these are properly set up before starting the work.

Going through each of these steps prior to beginning the work will improve the chances that the entire process will go smoothly. Once this preparation is complete you're now ready to begin the audit.

Additional SEO Auditing Tools

While your analytics package and Google Search Console both play a large role in your SEO audit, there are many other tools that can be an invaluable aid. A comprehensive discussion and review of these is provided in Chapter 3, but what follows is a short summary of the top tools you can leverage in an audit.

Crawlers

There are many high-quality crawlers available. These include (listed alphabetically):

- Botify provides both crawling and log file analysis solutions, as well as site change alerts and keyword tracking. In addition, Botify promotes their solutions as solutions tailored to specific market segments such as E-Commerce, Travel, Publisher, Classifieds, and Consumer Products. Botify is known for being well suited to crawling large scale web sites and can crawl 100 URLs requiring

JavaScript rendering per second. Botify pricing is by quotation only.

- Deepcrawl offers a solution that includes a crawler, a log file analyzer, and an additional tool called Deepcrawl Detect that is used to validate new code releases before you push them live. It also includes SEO Monitor Hub that can be used to detect changes to your site. Deepcrawl also supports crawls in excess of 10M pages. Deepcrawl pricing is by quotation only.

- Oncrawl also offers a web crawler and a log file analysis solution as well. Like Botify, Oncrawl discusses how their solution applies to specific industry segments. They also offer BI solutions and machine learning solutions to enable doing more with their data. Oncrawl was acquired by Brightedge in March 2022. Pricing starts at under $100 per month.

- Screaming Frog is highly popular partly because it provides a robust free version that can crawl up to 500 URLs, and the paid version is comparatively inexpensive (less than $200 per year). A separate log file analyzer is also available, also costing less than $200 per year. Screaming Frog requires you to provide your own computers to run their programs.

SEO Platforms

A wide range of SEO platforms are available. These include (listed alphabetically):

- ahrefs
- Brightedge
- Conductor
- Moz
- Searchmetrtics

- Semrush

- SEOClarity

Each of these provide a variety of ways to audit your web site. Figure 8-1 shows a sample of the data you can get from the Semrush SEO auditing tool.
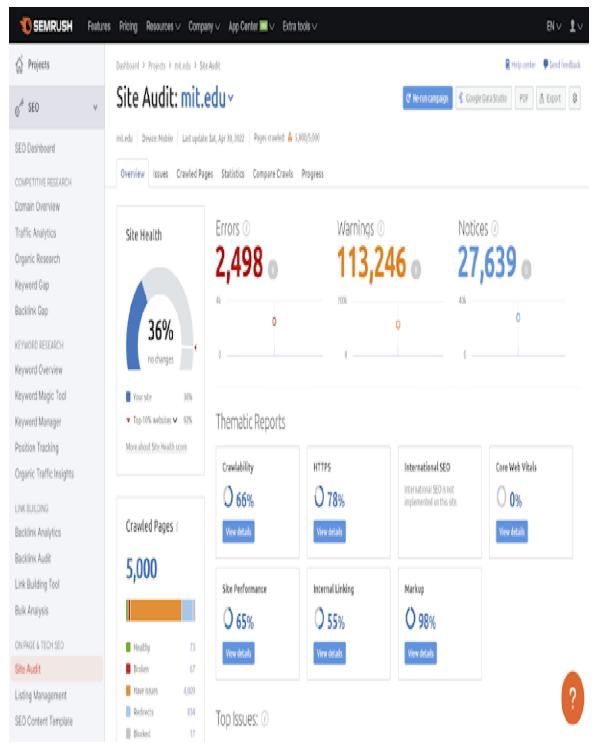
Figure 8-1. Sample Semrush Site Audit Report

The Semrush report summary provides a series of Thematic reports and you can click the "View details" buttons to get more details on each item. These types of automated SEO checks can uncover many types of problems and

be the start of your overall audit. In many of these tools the number of pages they let you crawl are limited and there are also many other things you will want to check so use this as the start of your audit – not the whole audit.

Core SEO Audit Process Summary

Before you start your audit you should put together a checklist of all the types of checks you want to perform and the issues that you want to look for. These are described in Sample SEO Audit Checklist. That SEO audit checklist can include six major areas of activity that you will want to execute:

1. Crawl the website with one of the crawlers discussed in SEO Auditing Tools. The crawl of your site is one of the richest sources of information available to you during an audit and crawls of large sites can take days so start your crawl at the very start of the audit process. You can see the types of checks that crawlers can help perform in Issues That Can Be Found Through Crawl Analysis.

2. Use Google Search Console to learn what you can on how Google sees the site. Bonus: Also use Bing Webmaster Tools as this may provide some different insights on how a search engine sees your website. You can learn more about how to use Search Console in our audit in Issues That Can be Found in Google Search Console.

3. Review your analytics data. Regardless of whether you use Google Analytics, Adobe Analytics or another platform, these should provide invaluable insight as to where the main level of activity takes place on your site. Analytics can help expose a number of other types of issues. You can learn more about those in Issues That Can be Found in Your Analytics. If you have not done so already setup a benchmark of current traffic and conversions, rankings, bounce rate and other factors. You will be able to use this in the future audits. If you have done this previously than leverage

it to see what has changed in terms of progress or setbacks for your site.

4. Use backlink tools to audit your backlinks. This can be used to help determine if your site has toxic backlinks or a weak backlink profile, broken incoming links and more. Details of the types of checks you perform with these tools are discussed in more detail in Issues That Can Be Found with Backlink Tools.

5. Have an expert SEO conduct a detailed human review of the site. While all the above tools are valuable sources of information there are many types of issues that are best spotted by the human eye. These are discussed further in Issues Best Found by Human Examination.

6. Review the Google SERPs to identify additional issues. This can include identifying missed opportunities and Structured Data errors, as well as if your title/description are not what you want. More information on how these checks can be used is contained in Issues Best Found by Reviewing the SERPs.

While these are the main areas of activity there are many other areas that you may choose to include in your audit. For example you may choose to perform a detailed review of other elements of your site such as Structured Data or hreflang tags. Or you have other areas where you know you have specific issues and you want to provide a more detailed examination of those. These are described in detail in Troubleshooting.

Sample SEO Audit Checklist

There are many different types of issues that you want to look for during a technical SEO audit. These can involve many different types of tools and approaches. As part of

Issues That Can Be Found Through Crawl Analysis

One of the critical steps in any audit is to perform a fresh crawl of your site, as well as an analysis of the log files for your site. Figure 8-2 shows an example of a crawl report from Semrush. Crawls like these can expose many core issues including the following:

Figure 8-2. Sample Semrush Crawl Report

- Incorrect or missing canonical tags. This can come in the form of links to pages that don't exist or links to incorrect pages. You can read more about how these tags should be used in Chapter 6.

- Broken internal links. These are links that return a 404 or a 410. This can happen if the links were improperly implemented or when a page is removed from the site. You can see a sample of 404 response codes found by Screaming Frog in Figure 8-3.

1	Address	▾	Content ▾	Status C.▾	Status	▾	Indexab ▾	Indexab ▾	1
470	http://whereis.mit.edu/bin/		text/html;	404	Not Found		Non-Index	Client Error	
617	http://gsc.mit.edu/2013/07		text/html;	404	Not Found		Non-Index	Client Error	
696	https://computationalengin		text/html;	404	Not Found		Non-Index	Client Error	
1169	http://web.mit.edu/facilitie		text/html	404	Not Found		Non-Index	Client Error	
1286	https://eapsweb.mit.edu/fo		text/html;	404	Not Found		Non-Index	Client Error	
1514	http://web.mit.edu/cheme/		text/html	404	Not Found		Non-Index	Client Error	
1527	https://giving.mit.edu/giver		text/html;	404	Not Found		Non-Index	Client Error	
2292	http://socialmediahub.mit.(		text/html;	404	Not Found		Non-Index	Client Error	
2402	http://kb.mit.edu/confluen		text/html;(	404	Not Found		Non-Index	Client Error	
2864	https://firstyear.mit.edu/fir		text/html;	404	Not Found		Non-Index	Client Error	
3565	http://scripts.mit.edu/~pale		text/html;	404	Not Found		Non-Index	Client Error	
3589	https://oge.mit.edu/finance		text/html;	404	Not Found		Non-Index	Client Error	

Figure 8-3. Sample Screaming Frog 404 Errors Report

- Use of NoFollow, Sponsored, or UGC attributes on internal links. As noted in Chapter 6, there is no reason to apply any of these attributes on links to other pages on your site.

- Pages blocked in robots.txt (a common error is blocking crawling of CSS or JavaScript files). Your audit should validate that you're not blocking crawling of any pages or resources that you want Google to be able to see.

- Bad redirects. Any of the well-known crawling tools will provide a report on redirects found on the site, including those that do not return a 301 HTTP status code.

- Missing title or heading tags. Each of the popular SEO crawlers includes reporting on title tags and heading tags that makes it easy to identify pages that are missing either of these tags. Figure 8-4 shows an example of missing H1 tags discovered in a crawl by Screaming Frog.

Figure 8-4. Sample Screaming Frog Missing H1s Report

- Duplicate title or duplicate H1 heading tags. Every page on your site should have title tags and heading tags that are unique to that page. If you are not able to come up with unique tags for a given page then you need to consider whether or not the page in question should exist. These issues are easily identified in your crawl report.

Note: if you have duplicate title and heading tags across paginated pages this is likely not a problem though we might still advise you include simple differentiators, such as a page number.

- Missing or generic meta descriptions. Instances of missing or generic meta descriptions should be validated. On very large sites it may not be possible to tailor meta descriptions for every page. In such a case, a more generic meta description or no meta description can be a valid choice to make. Nonetheless, identifying critical pages where these can be improved is a key activity during an audit.

- Missing images. Any missing images should be identified and fixed. These will show up as 404 errors in your crawler reports.

- Missing/inaccurate image alt attributes. Image alt attributes play an important role in assisting search engines in understanding the content of an image.

- Non-descriptive image file names. These are also used to help confirm the content of an image. Naming your image file names in a descriptive way helps reinforce the content of the image.

- Pages on your site that are not being crawled by Google, or important pages that are crawled with low frequency. The easiest way to check this is to use Log File Analysis to compare the pages crawled by Googlebot with the list of pages on your site found with your SEO Crawler. How to determine what may be causing this is discussed more in Pages Not Being Crawled later in this chapter.

- Duplicate content. These are pages that are essentially copies of other pages on your site or of other pages on the web. Crawl reports can help you potentially identify duplicate content by a variety of means. For detection of literally duplicate pages, tools use either a content hash or a checksum for the page. In addition, you can use other techniques to detect pages that may have too

much content duplicated, such as looking at pages that have duplicate title tags.

- Sub-optimal anchor text. Review the anchor text for generic phrases like "Learn More" or other language that is not likely to be descriptive of the destination page, as well as repetitive/overused anchor text.

- Content on your site that has no links pointing to it ("orphan pages"). Three of the basic ways that you may find external links are: (1) find pages in your XML Sitemaps that are not found in your crawl of the site; (2) find pages identified in your analytics program that are not found in the crawl; (3) find pages shown in Google Search Console that are not shown in your crawl, and: (4) identifying pages on your site that appear in your log files but not in the list of pages from a complete crawl of your site.

- Content on your site that is too many clicks away from the home page. The reason that this is important is that the number of clicks that your content is away from your homepage is an indicator of how important you see that content is to your site. If it's 5 or more clicks away then you're sending Google a strong signal that it's very low priority. Use this analysis to verify that you're comfortable with that for any pages for which that is the case.

- Over-reliance on PDF file content. Review the list of pages found on your site to see how many of them are PDF files. Google does not value PDF file content as much as HTML content. If you wish to offer users downloadable PDFs make sure that you also have an HTML version of the page.

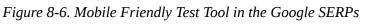## Issues That Can be Found in Google Search Console

Google Search Console is an invaluable tool for obtaining data on how Google sees your site. While this section will focus on Search Console you should also strongly consider setting up Bing Webmaster Tools as well as this will provide a different search engine's perspective on your site and in

some cases the information provided is different. Here are some of the types of things that can be found with Search Console:

- Pages on your site that are not being indexed by Google.

- Identify the top performing queries sending traffic to your site.

- JavaScript not rendering properly, causing content or links to be invisible to Google. An example of this is shown in Figure 8-5.

URL Inspection

Tested on: Mar 24, 2022, 9:27 AM

URL is available to Google

If it gets indexed and selected as canonical, it could appear in Google Search results with all relevant enhancements. Learn more

VIEW TESTED PAGE   Page changed? REQUEST INDEXING

| ✓ Availability | URL can be indexed | ˅ |

Enhancements

| ✓ Mobile Usability | Page is mobile friendly | › |
| ✓ Logos | 1 valid item detected | › |

Figure 8-5. Search Console URL Inspector

- Missing XML sitemap, or XML sitemap not specified in robots.txt. Sitemap.xml for multiple domains (e.g. international domains). Sitemap.xml with too many errors.

- Submit new or updated pages for re-crawling.

- Structured Data problems. Search Console offers high-level summary reports of structured data issues and also lets you drill down to see the details for each page. We describe how to do this in more detail in Validating Structured Data later in this Chapter.

- Page Experience issues. As discussed in Chapter 6, the Page Experience has many components. These include:

  - Lack of mobile-friendliness. Check your pages in the mobile-friendly test tool. You can also check this directly within the Google SERPs if you search on "google search console mobile friendly test." You can see an example of this in Figure 8-6.

Figure 8-6. Mobile Friendly Test Tool in the Google SERPs

- Presence of Insecure (http:) Pages. The presence of http: pages on your site is a negative ranking factor. Your crawl report will make identification of these pages easy to do.

- Use of interstitials on the site. Using interstitials on your site, particularly on initial page load, is treated by Google as a negative ranking factor. While this can be effective in generating revenue you should seriously consider the

negative implications, both in terms of organic search ranking and how users experience your site.

- Core Web Vitals factors such as slow page load speed, or high Cumulative Layout Shift (CLS). While page speed and CLS are small ranking factors they do matter. Search Console provides detailed information on the pages with errors on your site. You can see an example report in Figure 8-7.

Figure 8-7. Search Console Core Web Vitals Report

- See snapshot for some of the external links pointing to your site. This list is not comprehensive but can still be useful to see.

- Review and learn what you need to do to improve internal links on your site.

- Identify issues with the implementation of AMP on your site (if you have done this).

Issues That Can be Found in Your Analytics

Your analytics program also has a key role to play in your SEO audits. Whether you use Google Analytics, Adobe Analytics or some other analytics package you should be able to see the following information about your site:

- Overall traffic trends over time. Here you can see if your traffic is growing, shrinking or staying the same. For example, if you're traffic is down then you'll need to focus your audit on trying to find out why.

- Map what pages/site sections of your site are getting traffic. Some of the questions you might ask here include:

  - Are there important sections of your site that aren't performing the way you would hope?

  - Or are their specific key pages of your site that are underperforming?

  - These questions can lead you to focus your audit on these pages or site sections to figure out why they may be underperforming. Are there ways that the content can be improved? Does the site section / page have sufficient links from other pages on the site? Are there any external links to the page or section? How do your pages compare to those of competitors?

- Pages that have seen sudden traffic drops. Perhaps they have been penalized, a change to page content, or a recent change made to the site caused the search rankings to drop.

Issues That Can Be Found with Backlink Tools

As discussed in Chapter 10, links continue to play a major role in search rankings. For that reason understanding how your backlink profile compares to competition or any recent changes to that profile can be very revealing. There are many tools that can help with this. Some of the most popular tools are (listed alphabetically):

- ahrefs (https://ahrefs.com/)

- Majestic SEO (https://majestic.com/)

- Moz (https://moz.com/)

- Semrush (https://www.Semrush.com/)

The kinds of issues these tools can help you find include:

Weak backlink profile

Do your competitors have a much stronger link profile than you do? If their overall link authority is significantly greater than this can impede your ability to compete and grow traffic over time.

Backlink profile gaps

Are there certain types of links that your site lacks? Perhaps your competitors have earned links from many different types of sites: For example, media sites, blogs, education sites, government sites, and industry conference sites. In contrast it may be that your site only has links from blogs and industry conference sites. This type of coverage gap could lead to diminished overall authority.

Broken external links

External links that go to pages that return 40x errors do not add value to your sites ranking. For that reason, rapidly identifying those links and taking steps to repair them can be a fruitful activity. You can see more about how to do this in Chapter 10.

Link Disavowal

Many of these tools have the ability to create a link disavowal file or list of URLs to disavow. However, use this capability with care, as Google primarily just discounts toxic links so disavowing links should not be necessary. In addition, chances are that some of the links you submit for disavowal may be perfectly fine, so disavowing them would be harmful to your site.

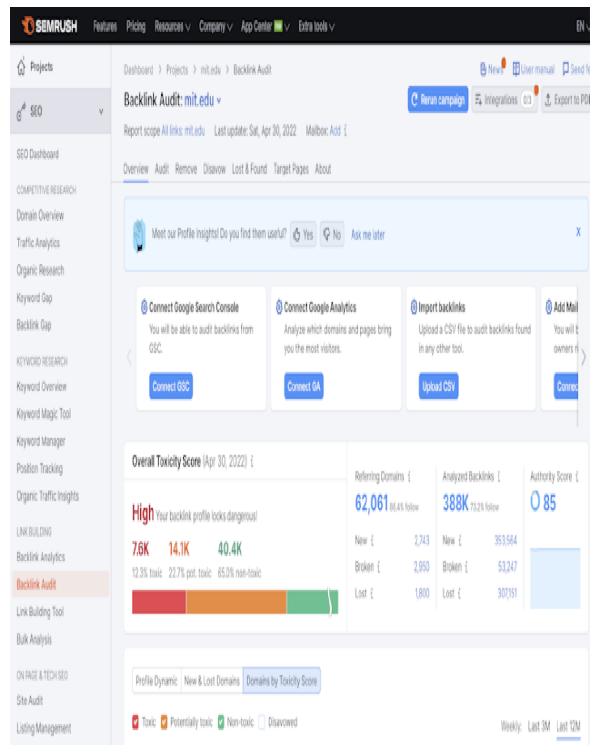You can see example output from the Semrush backlink audit tool in Figure 8-8.

Figure 8-8. Sample Semrush Backlink Audit Report

There are many other ways that backlink tools can be used to bolster your organic search visibility. These are discussed in more detail in Chapter 10.

Issues Best Found by Human Examination

Some issues are more easily recognized by human examination. Examples of these include:

Poorly structured site navigation

> Your site navigation plays a large role in how users navigate through your site. For that reason, it also plays a key role in how Google can access the site and how it sees what content you are prioritizing.

Opportunities for interlinking

> How you interlink on your site is also a powerful signal for SEO. This includes your overall navigation but also in-line interlinking within your content. It's often productive to review your content in detail and find opportunities to add more highly relevant in-context links in the content. Some of the crawling tools such as Oncrawl and Botify can perform an internal linking analysis and cross reference that with traffic and log file data.

Thin or poor-quality content

> While data from a crawl of the web site can identify pages with less content on them in many cases it's best to examine the pages by hand prior to concluding that they are thin content. Removing these pages from the Google index can in some cases offer powerful SEO upside, especially on sites that Google considers Your Money or Your Life (YMYL) sites.

Content created without sufficient expertise

> Content quality is a key part of long term SEO success, and determining the quality of content on a site is very difficult to do programmatically. Having someone with subject matter expertise review the content for quality and then addressing any shortcomings can offer significant benefits.

Content created primarily for search engines

Content created with the primary goal of ranking in search engines rather than for users is problematic. Detecting this requires evaluating content to see if your site is providing any unique expertise or perspective that other sites do not. It can also be helpful to review your organization's perspective during the content creation process. Was the primary goal to include certain keywords in the content or were you focused on helping users during the content creation process?

Problems with Experience, Expertise, Authoritativeness, and Trustworthiness (EEAT)

The concept of Expertise, Authoritativeness, and Trustworthiness (EAT) was first revealed by Google in their Search Quality Raters Guidelines (SQRG) in 2014. Since then Danny Sullivan confirmed in October 2019 that Google uses some signals to estimate the EAT level for a site, and that these are a ranking factor. Then this concept was expanded by adding a second "E" for Experience in December 2022. Evaluating how your site encourages EEAT and how others see your site has many subtleties that are best evaluated by a human, and it's an important part of SEO today.

The above list is by no means comprehensive as there are many different types of issues that can crop up in your SEO program where a detailed review by an experienced SEO is needed to finalize what actions should be taken with respect to potential SEO problems or opportunities.

Issues Best Found by Reviewing the SERPs

It can also be quite helpful to review the Search Results Pages (SERPs) on which your pages appear as well as the search results where you would like them to appear. At the start of your audit make a list of the search queries that you want to review. Some potential inputs to this list are:

- Make a list of the most important search queries to your site based on which ones deliver the most organic search traffic to your site.

- Make a list of the most important target search queries that don't currently deliver the amount of search traffic that you would like to be receiving.

- If you have an SEO enterprise platform (Brightedge, Conductor, Searchmetrics, Semrush, SEO Clarity) you can use that to find major changes in traffic on search queries over time. Each of these platforms will also show you which search result pages currently have which search features.

There are a number of different areas that you can investigate during your reviews of the SERPs:

Changes in Search page layout

Changes in the search page, including the addition or removal of search features, can have a large impact on the amount of search traffic you can receive. Review how these may have impacted your organic search traffic.

Titles and Meta Description that Google uses for your search listings

For those queries where one or more of your pages rank, review what Google is using for your page title and meta description. Are they the same as the title tag and the meta description that you wrote for the page? If not, consider why it is that Google may have chosen to write something different.

The main reason they rewrite these is that they feel that their version is a better fit to the search query. Based on that, consider how you might update your title or meta description. You may be able to create a title or meta description that covers the value proposition of your site better but still captures what it was that Google was looking for to describe your page.

Auditing Backlinks

SEO auditing can sometimes include auditing the backlinks pointing to your site. One of the key reasons for doing this is to determine if you have problematic links pointing to your site. You can learn more specifics about how to perform this analysis in Bad or Toxic External Links.

The first step in this type of audit process is to get as much information as possible on your backlinks. The most popular tools for pulling backlink data are:

- ahrefs (https://ahrefs.com/)

- Majestic SEO (https://majestic.com/)

- Moz (https://moz.com/products/pro/link-research)

- Semrush (https://www.Semrush.com/)

Each of these tools provides access to a significant link database. However, crawling the entire web is a massive endeavor and even Google doesn't crawl all of it. Google has the world's largest infrastructure for crawling which is certainly far larger than any of the comparatively small software companies listed above. Nonetheless pulling a list of links from any of them will provide a great deal of value in your backlink audit.

If you are able to afford working with one more provider of links, it is recommended that you pull link data from more than one vendor. If you're able to do this you will need to dedupe the list of links you pull from the two vendors. There will be significant overlap but it's also likely that your deduped link list will be as much as 50% larger than the list you would get by working with only one backlink vendor.

Once you have pulled the list there are two types of audits you can perform:

- Review your backlink to look for bad links as discussed more in Bad or Toxic External Links.

- Compare your backlinks with those of competitors. You can learn more about how to do this in Chapter 10.

SEO Content Auditing

Content plays a critical role in how much you can scale your SEO program. Once you have resolved your technical SEO issues working on the depth, breadth, and quality of your content is how you build scale. For that reason auditing your content is something that you should also do on a regular basis. The key states in a content audit are as follows:

1. Compile a complete list of all the content on your site. If you don't have an easy way to do this you can leverage a crawl of your web site to get a complete inventory of all your pages.

2. Build a list of pages that you want to review in detail.

3. Segregate those into classes:

    - Informational content targeted at the middle and the top of the funnel

    - Commercial and transactional content targeted at driving conversions

4. Conduct audits for each of these two different groups as further described in the next two sections.

Reviewing Content for SEO

One of the basic areas to investigate for all your content is how well it has been optimized for SEO. The first step is to ensure that your content has been created, or at least reviewed, by subject matter experts. It's not possible to build authority on a topic with content created by copywriters who are not knowledgeable about that topic. Imagine that other subject matter experts are reviewing your site content. What will they think about the content? Will they be impressed?

Then consider how well optimized the content is for SEO. Does the title tag include the main keyword related to the topic of the content? Does the content naturally use related keywords and synonyms that indicate that it has thoroughly covered the topic? Does the content link to other related

content on your site or other sites? Are you linking to/citing sources that support your claim?

Note that subject matter expert-created content will naturally be rich in keywords and ideally already be in good shape from an SEO perspective because of that. However, a detailed review by an experienced SEO may reveal some ways to tweak the title tags, meta description, or the content to help it rank higher.

Informational Content Audits

Informational content can play a large role in building your overall site authority and also fill the top of your purchase funnel with long term prospects. Your audit of your informational content should start with validating SEO optimization as described in Reviewing Content for SEO. Another big key to success with your informational content are these factors:

The breadth of your content

> If your coverage of the information people are looking for is too narrow (for example, you only create content for head terms) then most users will not be fully satisfied by the content they find on your site.

The depth of your content

> As with breadth, depth also helps you satisfy a larger percentage of the visitors to your site. The deeper you go on a topic the easier it is for you to be seen as authoritative.

When you consider the depth and breadth of your content it's helpful to build out a map of the topics covered by your competition. Crawl their sites if you can or use other means such as manual reviews to build a complete map of all the informational content that you can find on their sites.

Then ask yourself if they have significantly more informational content than you do on your site. Also, how does the content quality compare? Learn what you can about what you can do to improve the competitiveness

of your content. Having better quality content and more depth and breadth is a great place to start.

If you don't have the resources to create a broader and deeper array of content than one or more of your competitors consider doing so on some specific subtopics. As an output of this part of the audit, build out a map of the additional pages of content you want to create and/or pages that you need to enhance. This type of targeted effort can help you increase your market share around these subtopics and can potentially provide high ROI.

Commercial Content Audits

Commercial content pages often have significantly less text content than informational pages as much of the focus is on driving conversion. However, optimization of the content on the pages can still have a large impact. This starts with optimizing the title tags and the meta descriptions and other optimizations as described in Reviewing Content for SEO.

You can also consider adding content to the page that assists users in their experience with your commercial pages. Informational content about product/service options that is integrated with the page experience can help with conversion. Don't fall into the trap of sticking text content blocks at the bottom of your page where no one will ever read it, but instead find ways to integrate it into the overall experience of the page.

Another issue that is common on e-commerce sites occurs when they resell products from third parties and then use descriptive copy from the product's manufacturer. The problems arises because the manufacturer will provide that same descriptive copy to anyone that sells their product. As a result, there is nothing unique about it and is essentially duplicate content. If you can, write your own copy for the products and if there are too many of them find ways to add value (add user reviews or other value add content) or at least create your own content for the most important products.

Content Audit Summary

Creating the right mix of informational and commercial content can provide your site with an excellent way to grow market share, particularly if this is

combined with efforts to raise your brand visibility and attract high authority links to the site. You can read more about attracting links in Chapter 10.

There are also many other aspects of content that you can audit as well. These include following brand guidelines, how well it addresses your target personas, and its ability to help drive conversion. Describing these is beyond the scope of this book but remember to include those considerations during any content auditing process.

Troubleshooting

Even when you try to do all the right things within your SEO program you can still run into problems from time to time. This section reviews many common problem SEO scenarios and how to figure out where they are on your site.

Pages Not Being Crawled

As a first step in this process you can determine whether or not a page is being crawled by looking for it in your log files. If the page can't be found in your log files over an extended period of time, such as 60 days, there is a possibility that it is not being crawled at all. At this point you may need to consider the possibility that the page has some level of problem either preventing it from being crawled or causing Google to not want to crawl it. To investigate this further there are a series of checks you can perform to identify potential causes.

Blocked by robots.txt

Check your *robots.txt* file (http://www.yourdomain.com/robots.txt) to see whether you are preventing the crawlers from accessing parts of the site that you actually want them to see. This mistake is quite common.

Both Google Search Console (see Figure 8-9 and Figure 8-10) and Bing Webmaster Tools provide simple ways for you to see whether they are

aware of content that *robots.txt* is blocking them from crawling.
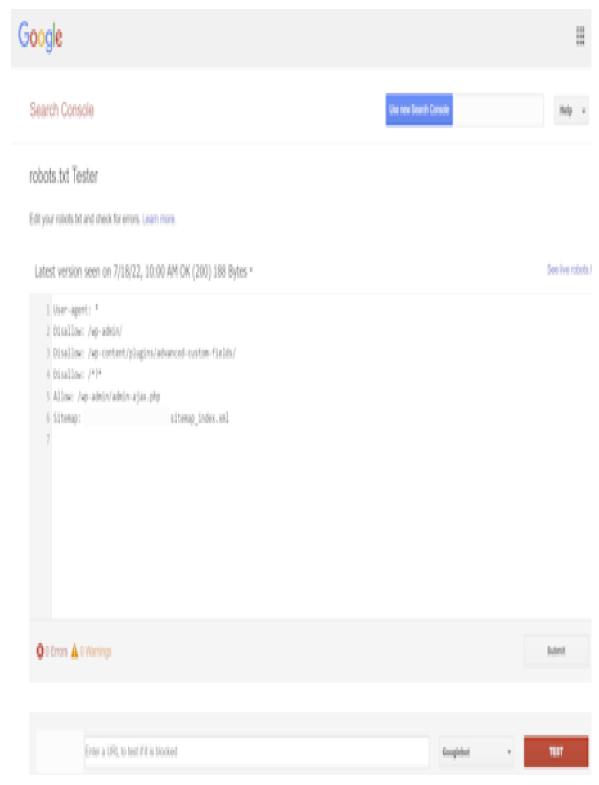
Figure 8-9. Google Search Console: restricted by robots.txt

Crawl requests breakdown

By response

OK (200)	98%	
Not modified (304)	1%	
Moved permanently (301)	< 1%	
Not found (404)	< 1%	
Page could not be reached	< 1%	

Rows per page: 5 ▼   1-5 of 5   〈 〉

Figure 8-10. Google Search Console: crawl errors

These reports are helpful when you have content on the site that has links to it (either internal or external) but that the search engines don't crawl because they are excluded from it in *robots.txt*.

The solution is simple: figure out what line in your *robots.txt* file is blocking the search engines and remove it

or update it so that the content is no longer being blocked.

Blocked by the robots meta tag

The robots meta tag in a page's header might look something like this:

```
<meta name="robots" content="noindex, nofollow">
```

As we discussed in "Content Delivery and Search Spider Control", a setting of noindex will tell the search engine that it is not allowed to include the page in its index. Clearly, you should check to see whether you have made this error if you find that the engines are not crawling certain pages that you want crawled. While the NoIndex tag does not explicitly tell the search engines to not crawl the page, after some period of time Google will stop crawling a page that is marked with this tag. Also, pages that have a noindex tag for a long period of time will have an implied nofollow tag to those pages.

A nofollow tag will tell search engines not to pass any link authority to the pages linked to on that page, however, search engines treat this as a hint (or suggestion), not a directive. If all the links on your site to a particular piece of content are nofollowed, you are passing no link authority to the page. This tells the search engines that you don't value the page, and as a result they won't treat the links as endorsements for it. While this does not specifically instruct the search engines to not crawl it or not index it, it can result in their choosing to not to do so.

Solving this problem requires locating the places where these robots meta tags are on your site and removing them. Note that the default setting for the robots meta tag is "index,follow", so there is no need to implement the

tag if that is your desired setting. Just make sure you don't have robots meta tags in place that change the default in places where that is not the desired behavior.

You can see if a page has no tag on it by checking the Coverage Report in Search Console, find the page in the crawl report from your web crawling tool, or looking in the source code for the page and searching for the "NoIndex" tag.

No direct links

You may find that a particular piece of content has no links to it. You can also make links invisible to the search engines (possibly unintentionally) by encrypting the links to the content in some fashion. If Google is aware of the page (perhaps via your XML sitemap or IndexNow for Bing and Yandex) it may still choose to crawl the page but that is not guaranteed.

The solution here is to make sure you implement plain-text (or image) links to the content. Better still, get some third-party websites to link to the content as well.

Form submission requirement

Requiring a login or some other type of form submission to see content is another common cause of non-spidering. Search engines will not attempt to fill out forms to see what is behind them. The simplest solution is often to remove the requirement for the form if you want the search engines to index this content.

However, some sites sell content on a subscription basis (also referred to as being behind a "paywall"), and they will not want to offer their content for free. In October 2008, Google announced the Flexible Sampling program (https://developers.google.com/search/docs/advanced/appearance/flexible-sampling), which allows subscription-based sites to have their content crawled and indexed by Google, but still allows the publisher to require human visitors to subscribe to access the content. You can read more about Flexible Sampling in "Content Delivery and Search Spider Control".

Session IDs

Session IDs confuse search engine crawlers: every time the engines come to your site, they see a different page. For example, they may see http://www.yourdomain.com?Ses-sID=2143789 one time and http://www.yourdomain.com?SessID=2145394 the next. Even though your intent is to track the session of a particular user, and you think of these URLs as the same page, the search engine does not. You can read more about session IDs in Controlling Content with Cookies and Session IDs.

Not enough link authority to warrant crawling

Sometimes the non-spidering problem has nothing to do with the issues we just discussed. The search engines may see the page just fine, but there may not be enough link juice going to it to merit inclusion in their main indexes. This is more common than people think, and it happens because the search engines do not attempt to index all the world's web pages.

For example, content that Google perceives to be of low importance (i.e., content that doesn't have enough link authority, or is perceived to be duplicate content) will be excluded from the main index. Many years ago, this content may have been relegated to what Google called its "supplemental index," but in 2014 Google's John Mueller confirmed that for the purposes of treating pages differently, Google no longer has a supplemental index (http://bit.ly/jan_13_hangout).

Google wants to emphasize the more important pages on the Web and doesn't want the rate at which it delivers search results to be slowed down by pages that most people probably don't want to see.

Page Indexation Problems

It's also possible to have content that gets crawled by Google but that doesn't get indexed. There are a few ways that this can happen such as:

The Page has not been crawled by Google

You can read more about this in the section on Pages Not Being Crawled in this chapter.

The page is marked with a NoIndex metatag

You can read more about this in Blocked by the robots meta tag section in this chapter.

The page does not have enough quality content for search engines to want to index it

Search engines want to deliver quality content to their users. If there is little or not content on a page, or all the content is hidden behind a form so search engines can't see it, they may choose to not include that page in their index. This can also happen if the content is of poor quality. In such cases the fix is to increase the quality and/or quantity of the content on the page. While looking at your crawl report data on the page may provide hints as to the quantity of content on a page (based on the page size) this issue is best detected by human examination.

The page does not have enough link authority for Google to want to index it

As discussed in Issues That Can Be Found Through Crawl Analysis there are a few ways that you can identify orphan pages on your site:

- Find pages in your XML Sitemaps that are not found in your crawl of the site

- Find pages identified in your analytics program that are not found in the crawl

- Find pages shown in Google Search Console that are not shown in your crawl

- Identify pages on your site that appear in your log files but not in the list of pages from a complete crawl of your site.

Even if there are no links from your own site to the page you can also check to see if there are links from third party sites to the page. That is

best done by using one or more backlinking tools to see if the page gets such links.

As a best practice, any page on your site you want the search engines to find should receive one or more links from other pages on your site. While search engines may choose to index pages that have no links (which they might still find if you list them in your XML sitemaps) it is far less likely that they will do so.

The page is subject to a manual action

Google does assign different types of manual actions to web sites or specific pages on websites. If your page is the subject of one of these actions this may be impacting its indexation. To determine if this is the case for your page, check the Manual Actions section within Search Console to see if you have received any such penalty. In order to address this issue you will need to understand the cause for the Manual Action and address the cause, and then file a Reconsideration within Search Console.

The page was removed by the URL Removal Tool in Search Console

You can manually remove pages from the Google search results using the URL Removal Tool. Note: these are not technically removed from the index. Once it has been removed by this tool it will be six months before it can be reincluded within the search results. You can cancel this removal at any time.

Duplicate Content

There are a number of techniques that you can use to attempt to detect duplicate content. These include:

- Identifying pages with duplicate titles and heading tags and manually checking them to see if they're duplicates

- Take a string of text from a page on your site and surround it in double quote characters and search on it in Google to see if it appears elsewhere on the web. Use a site: query to specify your domain to see if it is duplicated on your site. If the string your searching on is: "the boats circled endlessly in the storm", your query would be as shown here:

  ```
  "the boats circled endlessly in the storm"
  site:yourdomain.com
  ```

- Use Siteliner.com to examine your site and find duplicate content. This will provide a detailed report of duplicate content pages on your site. You can see an example of a report from Siteliner in Figure 8-11.

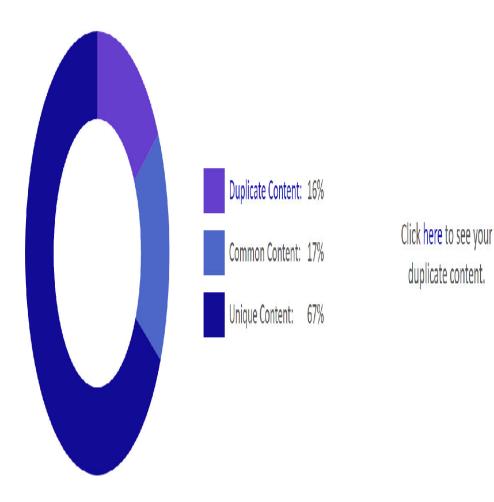Your Duplicate Content - Click below to see your duplicate content:

Duplicate Content: 16%

Common Content: 17%

Unique Content: 67%

Click here to see your duplicate content.

Figure 8-11. Siteliner Duplicate Content Report

- You can also check for content on your site being duplicated on other pages on the web using Copyscape. This can happen one of two ways:

    - Your content creators may have taken content from a third party web site and provided it to you as if it was their original work. Best practice is to proactively check for this

with any content provided to you prior to publishing, but you may also discover instances of this during an audit.

- Third parties may have scraped pages of your site and republished them on their site. You can address these instances by demanding that they take it down, notifying their web host of the problem as in many countries the hosting company is required to remove it once they are notified, and by also filing a DMCA takedown request with Google (https://support.google.com/legal/answer/3110420?hl=en).

NOTE

Important note: Quoting or citing content published on other websites in your content with appropriate attribution is not considered duplicate content, as long as the quoted content does not make up too large a portion of the page. Unfortunately, there is no clear metric for what constitutes "too large" but you should consider it a best practice that the great majority of content on a given page should be your own original content.

You can read more about detecting duplicate content here: https://www.searchenginejournal.com/seo-audit/duplicate-content/.

Broken XML Sitemaps

You may also have issues with your XML Sitemaps. These issues can cause search engines to ignore the content of your XML Sitemaps entirely and hence render them useless. The easiest way to determine if you have such a problem is to use the Sitemaps report in Google Search Console. This will provide you with a summary of all the errors which have been detected by Google. Also, most commercial crawling tools also provide means for crawling Sitemap files. These include (listed alphabetically):

- ahrefs

- Botify

- ContentKing

- Deepcrawl

- Oncrawl

- Screaming Frog

- Semrush

- Sitebulb

Ultimately, what you want is to have all the pages in the XML Sitemap meet the following criteria:

- The page returns a 200 HTTP status code (not a 3xx, 4xx, or 5xx status code) when loaded.

- The page does **not** contain a canonical tag pointing to a different URL.

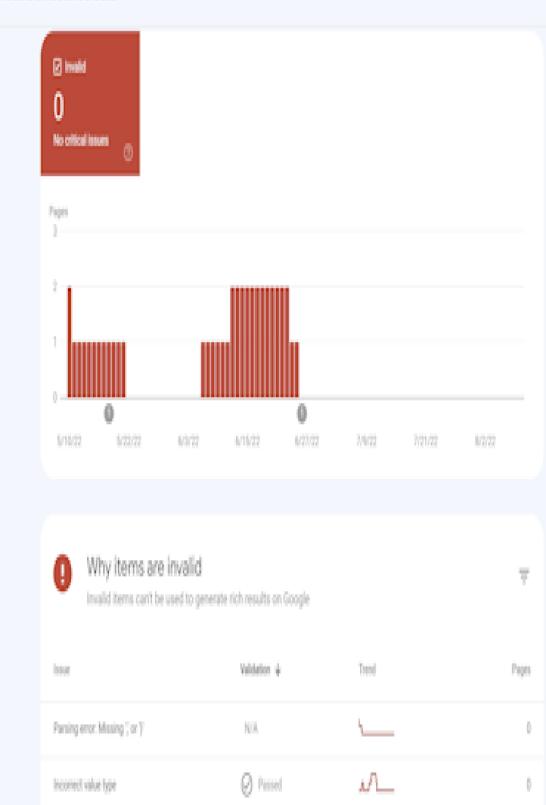- The page does not contain a NoIndex tag.

While it's not always possible to have all of your pages meet these criteria it's important that you get the number of pages that have one of these errors to represent less than one percent of the pages in the Sitemap file.

Validating Structured Data

Schema is a critical aspect of SEO and using it properly can result in enhanced display of your pages in the search results, and this can increase traffic to your site. You can read about this in detail in CSS and Semantic Markup in Chapter 6. There are many options for discovering or debugging Structured Data problems.

Search Console is a powerful tool for identifying where pages have problems and what the specific issues are on a page-by-page basis. You can start with the Unparsable Structured Data report to see how many total errors there are. You can see an example of this report in Figure 8-12.

Unparsable structured data

☑ Invalid
0
No critical issues ⊘

Pages

| 5/18/22 | 5/23/22 | 6/3/22 | 6/15/22 | 6/27/22 | 7/9/22 | 7/21/22 | 8/2/22 |

⊘ **Why items are invalid**

Invalid items can't be used to generate rich results on Google

Issue	Validation ↓	Trend	Pages
Parsing error: Missing ',' or '}'	N/A		0
Incorrect value type	⊘ Passed		0

From here you can click into each of the reported errors to get more details on what the specific issues are. Search Console also offers more specific reports for different types of Structured Data. These include:

- Breadcrumbs

- FAQ

- How-to

- Logos

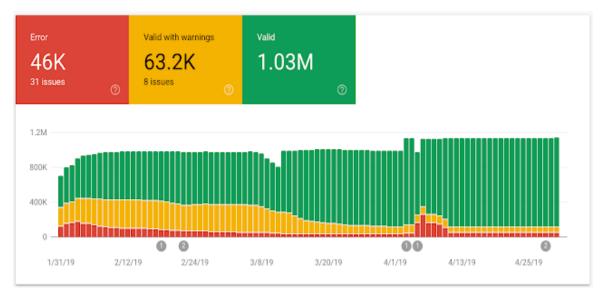- Products

- Sitelinks search box

- Videos

Figure 8-13 shows an example of what one of these reports look like.

Figure 8-13. Structured Data Errors Report

In this report you can click on the Error or Valid with Warnings summary boxes to get more details on the specific issues that were identified.

Search Console also provides the URL Inspection Tool. This provides the ability to validate if the page has any of these types of Structured Data on it:

- Breadcrumb

- Dataset

- Event

- FAQ

- Fact check

- Guided recipe

- How-to

- Image License

- Job posting

- Logo

- Product

- Q&A page

- Recipe

- Review snippet

- Sitelinks searchbox

- Special Announcement

- Video

The information provided includes descriptions of each item, and details about any warnings or errors found.

Another tool that is available is the Schema Markup Validator (https://validator.schema.org/). This tool allows you to input a single URL at a time to verify your schema implementation on that page. You can see an example of the Schema Vlidator showing a warning in Figure 8-14.
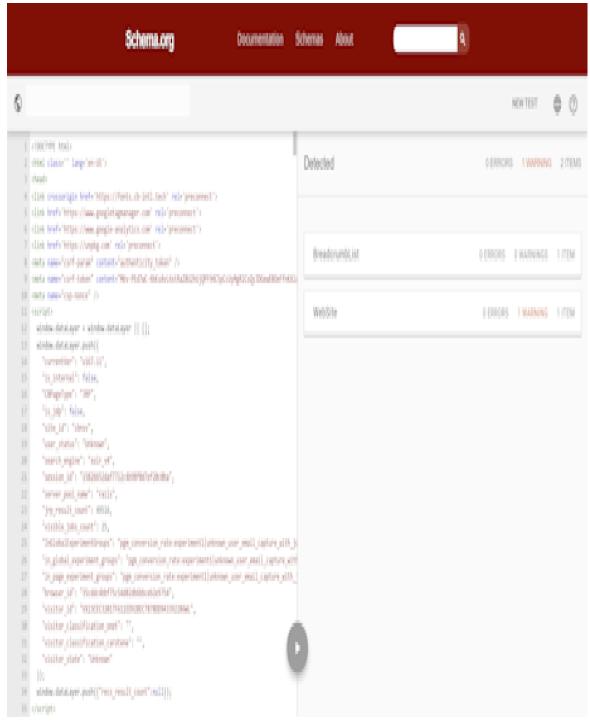
Figure 8-14. Sample Schema Validator Errors

In addition, you can also use the Rich Results Testing Tool
(https://search.google.com/test/rich-results). The specific purpose of this
tool is to determine whether or not your page is eligible for rich results in

the SERPs. You can see an example of what these results might look like in
Figure 8-15.

Figure 8-15. Rich Results Testing Tool Output

Validating hreflang tags

If you support multiple languages on your site then it's important to use hreflang tags to help search engines understand the relationships between the various versions of your international content. You will need to validate your implementation of the hreflang tags during your audit. But, if you determine that you have problems in ranking in alternative language markets one of the first areas to investigate is whether or not your hreflang tags are properly implemented. Proper implementation requires including tags that cross-reference all of the translations of the page and also include a tag that points to itself. This complete handshake ensures that Google will properly identify all of the translated versions of the page.

You can use tools such as Merkle's hreflang Tags Testing Tool (https://technicalseo.com/tools/hreflang/) or the hreflang Tag Checker Chrome Add in from Adapt Worldwide (https://chrome.google.com/webstore/detail/hreflang-tag-checker/hjgdcecfiohgajnhilmjhebdganpaomk?hl=en). You can read more about how to properly implement hreflang tags in Using Hreflang Annotations to Specify each Page and alternate versions' Languages and alternatively, Countries in Chapter 6.

Local Search Problems

For multi-location businesses local search plays a huge role in search visibility. Identifying whether or not you have a problem in local search. This can be done by a manual review of the search results but the complexity of this rises rapidly as the number of locations you have and the number of search queries you have scales up. In these scenarios using a tool that can review your listings in an automated way can be a big help. Selecting the tool that you want to use can be complicated because there are a large number of different local search tools available in the market. These include:

- Advice Local

- Birdeye

- BrightLocal

- GeoRanker

- Moz Local

- ReviewTrackers

- RioSEO

- Semrush Listing Management

- SOCi

- Surfer Local

- Synup

- Uberall

- Whitespark

- Yext

Yext is the tool that has the largest market share among medium to large enterprise organizations, but other tools have significant enterprise clients including: BrightLocal, Moz Local, RioSEO, and Semrush Listing Management. As Yext is the premium platform its cost is often higher than other platforms but it is a full service provider. For smaller businesses some of the other platforms may be a better fit.

Whichever platform you choose, use it to map out how your business is performing in the SERPs in order to identify the scope of any problem you may have there. You can also use the tool you choose to review the quality of your business listings in local search directories, business listing sites and Google My Business. Note that when you select your tool, review the quality of the data they provide you in these areas as it's a critical factor in selecting the tool that is best for your business. You can read more about Local Search and local search optimization in Chapter 11.

Missing Images

Image tags specify the URL where the image file can be found. However, it can happen that an incorrect URL gets placed in the tag or that the image file gets deleted or moved. Any of the commercial crawling tools identified in this chapter will identify any cases of missing images. Resolving the problem is straightforward as the options are:

1. Remove the image tag

2. Find the original image and update the tag to point to its current location or move it to the location pointed to by the image tag.

3. Find a new image, upload it to an appropriate location on the site and update the image tag to point to the new image.

Missing Alt Attributes on Images

Alt attributes play a key role in helping search engines better understand the content of an image containing the tag with the attribute. This is relatively easy to identify as any of the major commercial crawlers will identify any images missing these attributes. These include (listed alphabetically):

- ahrefs (https://ahrefs.com/)

- Botify (https://www.botify.com/)

- ContentKing (https://www.contentkingapp.com/)

- Deepcrawl (https://www.Deepcrawl.com/)

- Oncrawl (https://www.Oncrawl.com/)

- Screaming Frog (https://www.screamingfrog.co.uk/seo-spider/)

- Semrush (https://www.Semrush.com/)

Improper Redirects

While Google has confirmed that each type of 30x redirect you might use (e.g. 301, 302, 307, or 308) will still pass PageRank to the destination URL (https://searchengineland.com/google-no-pagerank-dilution-using-301-302-30x-redirects-anymore-254608) that doesn't mean that these redirects will pass all other signals through. For example, in the case of a 302 redirect Google may keep the source page containing the redirect in the indexFor example, in the case of a 302 redirect Google may keep the source page containing the redirect in the index, rather than the target page. As a result you may not have the page you want in the index..

For that reason, 301 redirects are still recommended as the preferred redirect for you to use. Finding out if you have been using any other types of redirects can be done by using the crawl report from whatever tool you have been using to crawl your site (examples: Botify, Deepcrawl, Oncrawl, Screaming Frog, Semrush). Each of these tools will provide you with a report where you can see any potential redirect problems.

Bad or Toxic External Links

How SEO professionals view link building has evolved significantly over the past two decades. It used to be common practice to buy links, implement large scale gust posting programs, and participate in large scale link acquisition schemes. Not everyone did these things of course but it happened enough and if your site has been around for a long time you may inherit some of these historical problems you may be stuck with the headache of having to clean it up.

Please note that most sites don't need to go through this type of process unless the scale of potentially bad links you have is extensive. If your site has just a few bad links Google will likely just mark them so they don't pass any PageRank. Before you commit to a large-scale project to clean up links, first make the determination that this level of effort is warranted. You can start by reviewing the Google guidelines on links. Three key resources for this are:

- https://developers.google.com/search/docs/advanced/guidelines/link-schemes

- https://developers.google.com/search/blog/2017/05/a-reminder-about-links-in-large-scale

- https://developers.google.com/search/docs/advanced/guidelines/paid-links

However, if you believe your organization has been systematically building links that violate Google's guidelines then you should implement a project to find the potentially offending backlinks and either get them removed from the pages containing them or disavowed using Google's link disavow tool (https://support.google.com/webmasters/answer/2648487?hl=en).

One of the first steps you should take is to talk to various members of your organization to see if you can discover any history of using link schemes to influence SEO. You should also consider the possibility that a more general marketing campaign may have been problematic.

For example, many large brands have participated in programs for paid guest post placements in which SEO was not a consideration. These may have been done simply to drive brand visibility, but if the placements were paid and links back to the web site were not marked as NoFollow then this could also present a problem.

Once you have conducted your investigation within your organization you should also pull the raw backlink data for your site and begin looking through those for more potential problem links. There are many tools in the market that you can help with this task, including:

- ahrefs (https://ahrefs.com/blog/fix-broken-links/

- Link Research Tools (https://www.linkresearchtools.com/seo-tools/dtox/)

- Majestic (https://majestic.com/guides/what-are-toxic-backlinks)

- Moz (https://moz.com/blog/bad-backlink-analysis-using-moz-link-explorer)

- Rmoov (https://www.rmoov.com/)

- Semrush (https://www.Semrush.com/blog/broken-link/)

Be aware that these tools can identify some of the links that may be problematic but you should not rely on them to identify all of the problem links. You will also want to review the full backlink files to see if you can identify problem links that you will want to address.

If you have a large quantity of backlinks this work might be quite extensive. However, you can simplify it somewhat by learning to look for certain types of patterns. For example, if you know that your organization has engaged in questionable link building practices one of the first things to look for is links that have anchor text that is almost too good.

Most links that your site receives will most likely use your domain name, organization name or relatively basic phrases links "click here." If you find that key pages on your site have rich anchor text in most of the links pointing to them that can be a leading indicator of a problem. For example, if you own a travel site and have a page for "Austin Hotels" and that page has 35 links and 28 of the links use "Austin Hotels" as the anchor text, there is a high likelihood that you have a problem.

A second pattern to look for is when deep URLs on your site appear to have too many external links pointing to them. For example, it's rare for e-commerce sites to have links to a large number of their specific product pages. If your backlink research determines that you have a material percentage of e-commerce pages that have external links, even if it's only one or two per page, and most of those also use rich anchor text, then that could be a key indicator of an issue.

In such a case you should dig deeper into those links. This should include asking around the organization to determine if there is any known history of any link acquisition programs related to those pages. You should also visit

the pages containing the links and see where they are on the linking pages for further indication as to whether or not the link was improperly obtained.

Another common problem is when your organization participated in large scale guest posting campaigns. If that has happened then focus on links that are from blogs on other sites. Some tips for that are to look for linking URLs that have one or more of these attributes:

- /blog/ is part of the URL

- The URL contains ?q= in the URL which is common in Wordpress blogs.

- Comes from a blog hosting domain such as wordpress.org, blogger.com, or other domains that host third party blogs.

- Use anchor text in the links that look like it may have been suggested by your organization.

Once you have completed your investigation you will need to decide if you want to disavow the bad links you may have found. If the quantity is less than ten percent of links to your site, and you also find that none of the key pages on your site don't appear to have links that are mostly bad, you may decide that you don't want to do so. However, if you find that there is a problem then you can use Google's link disavow tool to let Google know that you are disclaiming those links and you don't wish to receive and PageRank from them. You can read more about how to do that here: https://support.google.com/webmasters/answer/2648487?hl=en.

Single URL/Section Ranking / Traffic Loss

In the event that you detect a significant traffic loss to a single URL or a specific section of your site you'll want to try and determine the cause of the problem. Since the traffic loss appears to be localized on your site it's likely that the causes are somewhat localized too. This distinction should guide a key part of the plan for your investigation. Some key areas to explore are:

Tech changes

> Did your development team make changes to the impacted pages on the site? In particular, where these changes were made in a way that impacted only those pages that have shown a traffic loss. If so, review those changes and see if they may have hurt the SEO potential of those pages. For example, were pages marked as NoIndex or were canonical tags mis-applied to those pages on the site? Was that particular sub-directory or group of pages accidentally included as a "disallow" in the robots.txt file? Or were key portions of the content hidden from the search engines by requiring a user action to retrieve it? If you find these types of issues, then fix them and the cause of your traffic loss may be resolved.

Loss of internal links

> Another type of change that can impact SEO to a specific page or section of pages is if they have lost some or all of the internal links that previously existed to them from your site. Any of the popular crawling tools (Botify, Deepcrawl, Oncrawl, Screaming Frog or Semrush) will help you identify pages with lost internal links. Restoring those lost internal links may help address the traffic loss.

Loss of external links

> It's also possible that your pages may have lost external links that previously were helping them rank. This is where using the backlink tools (ahrefs, Majestic, Moz, or Semrush) can help you identify lost

links. However, be aware that these tools have some latency in that the external link may be gone for some time before the tools notice the lost link and show that in your reports. For that reason, make a point of identifying the most impactful external links to the impacted pages and checking the linking pages manually to see if you still have those links in place.

Unfortunately, there is not always that much you can do to address this particular issue. If you have lost some key external links you can always ask the site(s) that removed the links to put them back but they may simply decline. In such a case you either need to live with the traffic loss or launch a promotional campaign around those pages and hope to attract some links to them.

Changes to the content

Another potential driver of traffic loss could be any significant changes you may have made to the content of the impacted page(s). If material changes have been made to the content then perhaps Google finds the new version of the content less valuable to users and has lowered its ranking. If you have made significant changes to that content examine what was done and determine if it makes sense to restore some of the changed content to where it was previously. One resource for checking what changes have been made to your site is Archive.org (https://archive.org/), which is a site that crawls the web regularly and keeps historical snapshots of websites at various points in times.

Aging of the content

Ask yourself how current the content is on the page(s). Has it grown out of date? Are there events that have occurred that should be accounted for in the page content? A simple example for an e-commerce site would be that the version of the product you are showing is no longer the latest product.

Informational content can also become out of date too. For example, new scientific discoveries can render an explanation of how something

works inaccurate. Or dietary advice can change when medical advances cause us to understand more about how the human body works. The fix for this is to update the content and make it as current as possible.

In each of these cases you will need to make the changes and then wait for Google to recrawl those pages and process them through their algorithms to see if your fix was successful. Depending on the size of your site, the number of pages involved, and the frequency with which those pages are crawled you may need to wait anywhere from a week to several months.

Once you're certain that the pages have been recrawled you can determine if the fixes worked. If not you'll need to dive back in and consider what factors you may have missed could be causing your traffic loss.

Whole Site Ranking / Traffic Loss

Should your site experience a broad-based traffic loss some of the potential causes are similar to what you look for are similar to when you experience a drop to a single page or one section of a site. However, there are some differences and additional issues to consider.

Tech changes

Changes made by your development team can be a potential cause. In particular, were these changes made in a way that might have impacted most of the pages of the site? If so, review the changes and look for any potential SEO impact. For example, were pages marked as NoIndex or were canonical tags mis-applied across much of the site? Or were key portions of the content or navigation hidden from the search engines by requiring a user action to retrieve it? If you find these types of issues then fix them and the cause of your traffic loss may be resolved.

Changes in internal linking

Another type of change that can impact SEO across your site is significant changes to the internal linking on your site. Any of the popular crawling tools (Botify, Deepcrawl, Oncrawl, Screaming Frog or

Semrush) will help you identify changes in internal linking. Restoring the internal linking structure may help address the traffic loss.

Loss of external links

Have you lost key external links that used to point to your site? Backlinking tools (ahrefs, Majestic, Moz, or Semrush) can help you identify lost links. However, be aware that these tools have some latency in that the external link may be gone for some time before the tools notice the lost link and show that in your reports. For that reason, make a point of identifying the most impactful external links to the impacted pages and checking the linking pages manually to see if you still have those links in place.

Unfortunately, there is not always that much you can do to address this particular issue. If you have lost some key external links you can always ask the site(s) that removed the links to put them back but they may simply decline. Addressing this will require launching promotional campaigns to attract more high quality links to the site.

Changes to site content

Another potential driver of traffic loss could be any significant changes you may have made to site content. If material changes have been made to the content across the site, such as deciding to significantly reduce the amount of content, then Google may find your site less helpful to users and therefore has lowered its ranking. Consider reversing those changes if it's a probable cause for your traffic loss.

Aging of the content

How often do you update content on your site? Has it possibly become out of date? If so, consider investing the time to update it. Also, implement policies to update your content on a regular basis so that it does not get to be out of date in the future.

Google algorithm updates

Has there been a recent announced update to the Google algorithm? Consider that the update could be a core algorithm update that Google has confirmed or other updates that get no confirmation. Perhaps industry sources such as Search Engine Roundtable (from Barry Schwartz) and SEOs on Twitter are discussing industry buzz about some type of Google change.

Then, determine if the timing of this change lines up with when your traffic drop began. If so, then research all that you can about what other people are saying about the update. See if those factors might apply to your site and do what you can to address them.

Expertise, Authority, and Trust (EAT) issues

Google has a set of concepts Expertise, Authority, and Trust (EAT) as a set of factors that relate to the likelihood that the content on a site will provide value to users. I.e. was the content created by expertise, does your organization have authority on the topic, and can your motivations be trusted? These concepts were originally defined by Google in their Search Quality Raters Guide (SQRG) for use by human search quality reviewers. You can read more about EAT in Google's EAT and YMYL in Chapter 6.

As defined in the SQRG, Google does not use EAT as a ranking factor. However, in October of 2019 Google's Danny Sullivan confirmed that Google does look at other factors that help them determine "if content seems to match E-A-T as humans would assess it."

Most EAT changes you might make to the site would likely not cause major changes in traffic loss but there are certain specific types of changes that can have a larger impact. For example, if you have a site that deals with what Google refers to as Your Money or Your Life (YMYL) topics and you publish a lot of content of questionable quality.

Google is very sensitive about YMYL sites and if it believes that your site is providing bad or questionable information on YMYL topics it may well lower the rankings for your site in a broad way. In this case,

the solution is to lower remove the poor quality content or improve it significantly.

In each of these cases you will need to make the changes and then wait for Google to recrawl those pages and process them through their algorithms to see if your fix was successful. Depending on the size of your site, the number of pages involved, and the frequency with which those pages you may need to wait anywhere from a week to several months.

Once you're certain that the pages have been recrawled you can determine if the fixes worked. If not you'll need to dive back in and consider what factors you may have missed could be causing your traffic loss.

Page Experience Issues

Google began rolling out the Page Experience update in June, July, and August of 2021. Then in February of 2022 it was extended to the desktop. The Page Experience algorithm includes several different components:

Mobile- Friendly Site

Measures the quality of the user experience your site offers on mobile devices.

Use of Interstitials

Evaluates if/how you use interstitials in a way that impedes the user experience with your content.

Use of HTTPS

Lowers rankings for pages that are not secure (e.g. serve up as http pages)

Cumulative Layout Shift (CLS)

Measures how much shifting happens to your pages during page load.

First Input Delay (FID)

Calculates how long it takes a browser to respond to a user input request.

Largest Contentful Paint (LCP)

Measures how long it takes for the largest above the fold content elements to display when your pages load.

You can read more about the Page Experience signal in Chapter xx. Google Search Console offers an overall Page Experience Report which you can see here: https://support.google.com/webmasters/answer/10218333?hl=en. The 4th through 6th signals in the above list are also rolled up into a single concept known as Core Web Vitals (CWV). Search Console also offers a report specific to Core Web Vitals that you can see here: https://support.google.com/webmasters/answer/9205520?hl=en.

Detecting and resolving problems with mobile-friendliness, interstitials, or HTTP: pages is comparatively easy, as follows:

Mobile-Friendly Site

Google's Mobile-Friendly Test Tool (https://search.google.com/test/mobile-friendly) makes evaluating this quite easy. Sample URLs from each of your site's major page templates and test them in the tool to see how they score.

Use of Interstitials

What Google is looking for with interstitials is outlined in this post: https://developers.google.com/search/blog/2016/08/helping-users-easily-access-content-on. The easiest way for you to validate if any use you may be making of interstitials is a problem is to become familiar with the contents of this post and compare that with what you're doing. If there is a problem you may want to modify how you are using them.

Use of HTTPS

There are many methods available to find the HTTP: pages on your site:

- Review the output of your crawl report to see if any pages were found during the crawl that use HTTP: instead of HTTPS:

- Check your XML Sitemaps to see if any HTTP: pages are listed within

- Use Chrome to check if a site is secure as further described here: https://support.google.com/chrome/answer/95617.

- Create a Domain Property or a URL prefix property and follow the steps further described here: https://support.google.com/webmasters/answer/10218333?hl=en#find_http_urls.

If you find any problems with these three aspects of Page Experience, coming up with how you want to fix them is not necessarily simple. This may involve design decisions and/or business decisions that may be complicated to get agreement on within your organization. However, from a technical perspective the needed changes are usually relatively straightforward to define (though the your site platform may make it hard to implement).

Detecting problems with your Core Web Vitals (CWV) is also easy to do but determining how to resolve them can be a lot more complex. Numerous tools are available to help with detecting CWV issues, including:

Chrome User Experience Report (https://g.co/chromeuxdash)

From the Chrome team this is actually a Google Data Studio dashboard that you can use to see aggregated real world performance data for actual users on your site. You need to configure it to set it up but the data is invaluable as it also helps you understand the mix of different types of devices that people use to access your site. This is also referred to as the CrUX report. This report can also be seen within Page Speed Insights and Google Search Console.

Google Page Speed Insights (https://pagespeed.web.dev/)

Highly valuable as this tool is from Google and it will show you data from the CrUX report. It will show "lab data" based on real time testing performed by PageSpeed Insights if there is not enough CrUX data available. PageSpeed Insights also provides a list of potential items to fix.

Google Lighthouse

This tool is also from Google and is embedded within Chrome and can be accessed by clicking the three dots at the top right of the browser (next to your picture) then More Tools and then finally Developer Tools. From there you can use Lighthouse Tools to load the current page you're on and provide a full report of page performance.

GTMetrix (https://gtmetrix.com/)

GTMetrix also offers a well-known free tool for monitoring your page speed along with a variety of test centers so you can test performance from different locations. In addition, they offer (for a fee) the ability to set up ongoing metrics that will trigger alerts when page performance falls below a threshold that you set.

Web Page Test (https://www.webpagetest.org/)

This is another well-known third-party tool offering a wide variety of metrics and diagnostics. The site also has an active blog and highly active forums that are both rich with information on the tricky aspects of page speed optimization.

While there are many tools available to monitor the performance of your web site, truly speeding up your pages can be quite complex. Since interactions between your web server and the user's browser are multi-threaded it can be hard to determine where the actual bottlenecks are. In addition, some of the problems may not be in how you coded your page but could instead be within your overall hosting infrastructure. For example:

- You may need to upgrade the web server you are using. This could be moving from a shared server to a dedicated one, or if you already have a dedicated server you may need to upgrade to a faster one.

- You may simply need to get more memory for your hosting server.

- You may need to upgrade the connection bandwidth to your server.

- You may have to configure a CDN or change the configuration of your CDN.

- You may not have enabled gzip compression on your web server.

- You may need to update your database servers.

This by no means a complete list of potential issues and are just examples of the issues you may be facing. Further determining what mix of these types of changes may be needed to improve the performance of your page(s) can be very hard to do, and in some cases the changes can be very difficult to implement.

For many organizations working with someone who specializes at addressing these types of issues may be the best course of action. Smaller organizations may need to rely on a best practices approach or if you're using a platform like Wordpress there are also many plugins that you can use to help speed up your site.

Thin Content

A page is considered thin in content if it has little information of value to users. This is considered bad as sites with significant amounts of thin content on their site can have their rankings lowered and can even be penalized by Google. You can read more about thin content penalties here: https://developers.google.com/search/docs/advanced/guidelines/thin-content.

Thin content is not necessarily about the number of words on a page and can include these types of scenarios:

- Auto-generated content.

- Affiliate pages that simply repeat commonly available information.

- Doorway pages create for the sole purpose of ranking for search queries.

- Content scraped from another web site.

- Substantially similar pages that match highly similar search phrases.

Note that pages with very little text content might be considered thin content but there are many cases where this is not a sufficient measure. For example, e-commerce pages may have a number of product listings and only very basic product descriptions on them and still be considered a page that offers some user value.

Some of the key questions to consider when working to determine whether or not you have thin content on your site:

1. Have you been auto-generating content without taking sufficient care to ensure its quality? Note that advances in AI are making machine-generated content something that may soon be practical, but you need to anticipate that human reviews of the content (and ownership of the final output) are still a requirement.

2. If you're running a site that monetizes through affiliate programs, are you taking care to create unique content of high value to users?

3. Have you created a large number of pages for the sole purpose of ranking for a wide variety of search phrases with little attention to the content on those pages? These pages may be considered doorway pages and are often poorly integrated into the core part of your site.

4. Does your site contain content that has largely been scraped from other sites?

5. Do you have many pages designed to rank for minor variants of search phrases? For example, if you offer a site that assists users in creating resumes do you have separate pages for phrases like these:

- Resume writing services

- Resume writing service

- Services for writing results

- Resume writing assistance

Asking these types of questions can help you find pages that may be considered as thin content pages. Ultimately, it's very helpful to examine the pages by hand and evaluate the content and ask you whether or not they are actually helpful to users.

Poor Quality Content

While this may seem highly similar to thin content there are some different aspects to it. Examples of different types of poor-quality content include content which:

- Is low in relevance to your site

- Relies on keyword stuffing

- Is inaccurate or misleading

- Offers users one thing but then tries to sell them something else (bait and switch content)

- Does not address the user need in a meaningful way

One leading indicator of pages that may be of poor quality is whether or not it attracts any organic search traffic. However, that is not itself a clear indicator that a page is of poor quality. For example, pages may still offer

highly useful content on a topic that does not have high search volume, or it may have other reasons why it is not ranking.

As with thin content, one of the best approaches to identifying where you may have poor quality content is to ask some key questions:

1. Has your content been created by subject matter experts or at least reviewed by subject experts before publication?

2. Does all of the content fit the overall theme of the site?

3. Has all your content been fact checked for accuracy?

4. Do the pages on your site deliver content highly related to the title tag of their pages?

If the answer to any of the above questions is no then you should look at those pages to see if they contain content that would objectively be considered to be of poor quality. Read the pages carefully and evaluate how it helps users. Better still, even if it is decent or better content, take the opportunity to improve it while you're there!

Content Which is Not Helpful to Users

Google's primary goal in constructing their search results is to deliver pages to users that help address their needs. This helps them maintain their market share and increase usage of their search engine. However, ranking in search engines is of critical importance to many organizations and some of them implement programs to generate content solely for ranking in search engines. The problem with starting with this mindset is that it can lead to content spew – creating large volumes of content that does not add that any new value or perspective on the topic covered ("Unhelpful Content").

Google's Helpful Content Algorithm is designed to identify this type of content and impact the rankings of sites that follow this type of practice with their content. Here is what Google said about the algorithm in their blog post about it

(https://developers.google.com/search/blog/2022/08/helpful-content-update):

"To this end, we're launching what we're calling the "helpful content update" that's part of a broader effort to ensure people see more original, helpful content written by people, for people, in search results."

Unfortunately, the impact of this algorithm is sitewide. It doesn't just lower the rankings of the Unhelpful Content, it lowers rankings for all of the pages on the site containing that content. In this regard the Helpful Content Algorithm has a punitive aspect to it (though Google does not label it as a penalty as that term is applied by them only to manual actions). Further, resolving the problems with the Unhelpful Content doesn't result in speedy recovery. Google has indicated that once the problems are addressed it can take many months to recover your previous rankings across other parts of the site.

In evaluating your content to see if this algorithm is impacting you here are some questions that you can ask yourself:

1. Is the primary focus of your content to help users?

2. Does your site have a clear purpose or are you generating content in many areas to see what works?

3. Is your content created by people with clear expertise on the topic?

4. Was your content created by people rather than machines?

5. Does your content add new value to the web on the topic that it addresses?

6. Did you create the content simply because it was trending, or does it actually address user needs?

7. Are you creating large volumes of content on many different topics just to gain search rankings?

Ultimately, one of the best ways to determine if you have content which may be considered Unhelpful Content is in reviewing your motivations in

creating it and the approach used in the process. If the primary goal is to gain organic search traffic than that is a leading indicator of a potential problem. Some other aspects of your approach to content creation process that can be problematic are:

1. Using AI / machine learning algorithms to create the content.

2. Failure to use subject matter experts to author your content.

3. Providing SEO-centric instructions to guide the content creation. This includes specifying content length and keyword-centric guidelines in your instructions to the author.

In general, the best strategy to guide your content strategy is to focus on the unique user value that you're addressing with the content you create. However, if you're site is hit by the Helpful Content Algorithm your options for addressing it are:

1. Remove the Unhelpful Content from your site.

2. Add the noindex tag to the Unhelpful Content pages.

Removing the pages from the index via noindex tags should be viewed as a patch however as the best solution is to remove the offending pages from the site. Certainly if the pages were created primarily for search rankings and are not of quality as a result, they are not likely to be helpful in building your reputation and relationship with the users that encounter them.

Google Altering Your Title or Meta Description

As we've discussed in this book one key area of SEO is to create well-written page titles and meta descriptions as these are often what will be shown for your page when a listing is included for it within the SERPs. These are parameters that appear within the <head> section of your web page and typically are configured within your CMS or e-commerce platform. If you're using a CMA such as WordPress you may require a plugin such as Yoast SEO to set these values.

However, Google sometimes chooses to not use the title tag that you specified and instead configures their own. There are several reasons that Google may choose to do edit your title or meta description prior to showing them in the SERPs:

- They may not accurately reflect the content that can be found on your page.

- They may not accurately reflect how the content on your page offers value related to the specific search query entered in by the user.

- You may have been overly self-promotional in how you wrote those tags.

The most common way that you can discover when this is happening is by trying sample search queries and then seeing what Google shows for your page in the search results. If what Google is showing differs from what you created then the best approach to fixing the problem is to take a cut at rewriting the titles and/or meta descriptions that Google does not appear to like. Try to address the likely cause for why Google felt they needed to rewrite them.

However, first consider whether that is even necessary. Perhaps Google is only editing them for very specific user search queries (and not others) and their title / description may be better than yours as relates to that specific query. Your titles and meta descriptions should focus on the primary intent of the page and if Google is rewriting them for secondary intents that are not the core purpose of the page then perhaps you should leave well enough alone.

Hidden Content

In "Content Delivery and Search Spider Control", we discussed ways that you can hide content from the search engines when you want to. However, at times this is done unintentionally—that is, sometimes publishers produce

great content and then, for one reason or another, fail to expose that content to search engines.

Valuable content can be inadvertently hidden from the search engines, and occasionally, the engines can find hidden content and construe it as spam, whether that was your intent or not.

Identifying Content That Search Engines Don't See

How do you determine when you have unintended hidden content? Sometimes the situation is readily apparent; for example, if you have a site that receives high traffic volume and then your developer accidentally places noindex tags on every page on the site, you will see a catastrophic drop in traffic. Most likely this will set off a panicked investigation, during which you'll quickly identify the noindex issue as the culprit.

Does this really happen? Unfortunately, it does. Here is an example scenario. Suppose you work on site updates on a staging server. Because you don't want the search engines to discover this duplicate version of your site, you keep the pages on the staging server noindexed. Normally, when you move the site from the staging server to the live server, you remove the noindex tags, but unfortunately, many site owners forget to do this.

This type of problem can also emerge in another scenario. Some webmasters implement a *robots.txt* file that prohibits the crawling of their staging server website. If this file gets copied over when the site on the staging server is switched to the live server, the consequences will be just as bad as in the noindex scenario just outlined. The best way to prevent this type of situation is to implement a series of safety checks on the site that take place immediately after any update of the live server.

There are potential problems, however, that are much more difficult to detect. First, with a new site launch, you won't have any preexisting traffic, so there will be no drop in traffic levels to alert you that something is wrong. In another scenario, you may have an established site where you accidentally do something to hide only a portion of the site from the engines, so the issue is less obvious.

Regardless of your situation, web analytics can help you in the detection process. Use your analytics software to find pages on your site that get page views but no referring search traffic. By itself, this is not conclusive, but it provides a good clue as to where to start. Note that the converse of this is interesting for another situation—if you see content that is getting search referrals even though you don't want or expect it to, you may want to hide that content.

Another data point you can examine is the number of pages the search engines report as indexed for your site. In a new site scenario, you can look at this to see whether the search engines appear to be picking up your content. For example, if you have a site with 1,000 pages with a good inbound link profile, and after three months only 10 pages are indexed, that could be a clue that there is a technical problem. Using multiple sitemap files, one for each site content area covering a specific segment of URLs, can be helpful in diagnosing such problems.

You do need to be careful not to overreact to the count of indexed pages, because the numbers that the search engines report will naturally fluctuate quite a bit. But if you are aware of the types of numbers typically reported for your site, and they drop to an unusually low level and stay there (or keep dropping), you probably have a problem.

Identifying Hidden Content That May Be Viewed as Spam

Hidden text is one of the challenges that webmasters and search engines still face. Spammers continue to use hidden text to stuff keywords into their pages, for the purposes of artificially boosting their rankings. Search engines seek to figure out when spammers are doing this and then take appropriate action. There are many ways to create hidden text unintentionally, though, and no one wants to be penalized for something they did not intend to do. Google's Webmaster Guidelines for hidden text (http://bit.ly/hidden_links) state the following:

Hiding text or links in your content to manipulate Google's search rankings can be seen as deceptive and is a violation of Google's Webmaster Guidelines (https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines).

If you're using this technique to try and stuff keywords in your webpage you're definitely over the line and into blackhat territory. However, there are certainly scenarios where your CMS may create some hidden text as outlined in the next section.

Unintentionally creating hidden text

There are a few ways to create hidden text without intending to do so. One of the most common ways is via your CMS, which has some CSS-based methods built into it. For example, many content management systems use the display:none technique to implement drop-down menus or other widgets that "expand" to display more text when clicked. Tab folders are a great example of this. Sometimes the display:none technique is used in user-generated content systems where the page normally shows the number of comments on a post, but suppresses the text "0 Comments" in the event that no comments have been made.

People can also commonly create hidden text when they provide enhancements for the visually impaired. For example, you may have a short video on your web page and want to provide users with a text description of the content. You may not want to place the text on the page, as it might make the page look cluttered to a user with normal vision. The solution some people use to serve both audiences is to hide the text from the sighted users.

Many of these scenarios have no SEO value, even when manipulated by spammers. These types of techniques generally do not carry a risk of being penalized, because there is no reason to suspect negative intent.

Conclusion

Audits play a critical role in your SEO program. Even highly aware SEO organizations can have problems creep into their site, be impacted by Google algorithm changes, or discover new opportunities. A well-planned SEO auditing program can minimize the scope and risk of SEO problems as well as enable you to remain highly competitive for SEO.

About the Authors

Eric Enge is the founder and CEO of Stone Temple Consulting, a leading SEO firm that serves a wide variety of companies, including a number of Fortune 100 companies. He writes regular columns in Search Engine Land and Search Engine Watch, and has also provided feature articles to SEOmoz. The interview series he publishes on the Ramblings about SEO blog (http://www.stonetemple.com/blog) regularly provides information directly from senior people fro Google and Microsoft on a regular basis. Eric is also a highly regarded speaker, and regular speaks on Internet marketing topics at conferences such as Pubcon, Search Engine Strategies, Search Marketing Expo, the American Marketing Association, SEMPO, Internet Retailers, O'Reilly, Web 2.0, and many others. You can also see content published by Eric on a regular basis by following him on Twitter (@stonetemple) or putting him in one of your circles on Google+ (+Eric Enge).

Stephan Spencer is an internationally recognized SEO expert, Internet entrepreneur, sought-after professional speaker, and bestselling author. He is the author of Google Power Search and coauthor of The Art of SEO and of Social eCommerce, all published by O'Reilly. He is the host of the podcasts Marketing Speak and The Optimized Geek. Stephan produces and presents the 3-day training intensives Traffic Control and Passions into Profits. He is producing a video-based SEO training and coaching program at ScienceOfSEO.com. Stephan founded the interactive agency Netconcepts in 1995 and grew it into a multinational SEO firm, with dozens of staff in each of its U.S., New Zealand and China offices.

Jessie Stricchiola is the founder of Alchemist Media, Inc., a San Francisco search engine marketing company. She began her search engine optimization career in 1997. For the past five years, she has been a guest speaker at each of the search industry's leading conferences including Incisive Media's Search Engine Strategies Conferences, Danny Sullivan's Search Marketing Expo, WebmasterWorld's PubCon, ad:Tech, WebGuild, and Shop.org. As one of the original nine founders of SEMPO (the Search Engine Marketing Professional Organization), she served for two years on

the Board of Directors for the organization. Jessie has been interviewed by numerous trade publications and media outlets,including The New York Times, The Wall Street Journal, Wired Magazine, The Washington Post, SearchEngineWatch.com, CNET.com, CNBC, NPR, and The BBC. Her work on click fraud is featured in The Google Story by David Vise.