SEJ EBOOK

HOW SEARCH ENGINES WORK

WRITTEN BY
DAVE DAVIES

highervisibility

MORE TRAFFIC IS GOOD
MORE SALES ARE BETTER

Partner with an award winning digital marketing agency that has proven to deliver customers through SEO, PPC, and Social Media services.

MARKETING SOLUTIONS SPECIFICALLY CATERED TO YOUR BUSINESS

- SEO
- PPC
- Social Media
- Web Design
- SEO Audit
- Link Building

COMPANY RECOGNITIONS

888.212.4458 info@highervisibility.com **www.highervisibility.com**

CONTENT

CH 1

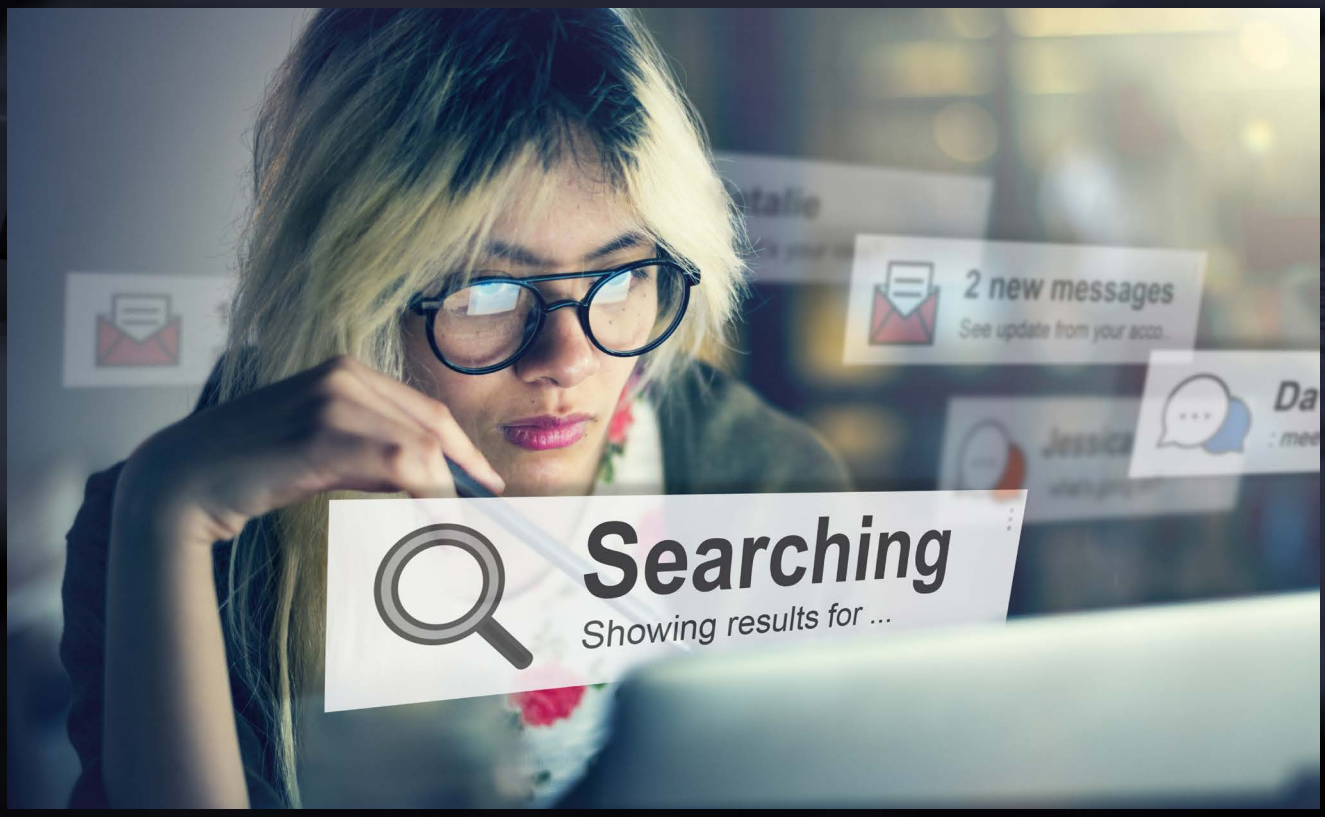
HOW SEARCH ENGINES CRAWL & INDEX

Optimizing websites without first understanding how search engines function is akin to publishing your great novel without first learning how to write.

Certainly, a thousand monkeys at typewriters will eventually create something useful (at least this monkey likes to think he does from time to time), but it's a lot easier if you know the core elements of a task beforehand.

So we must understand how search engines work to fully understand how to optimize for them.

While we will be focusing on organic search, we must first briefly talk about one critical truth about search engines.



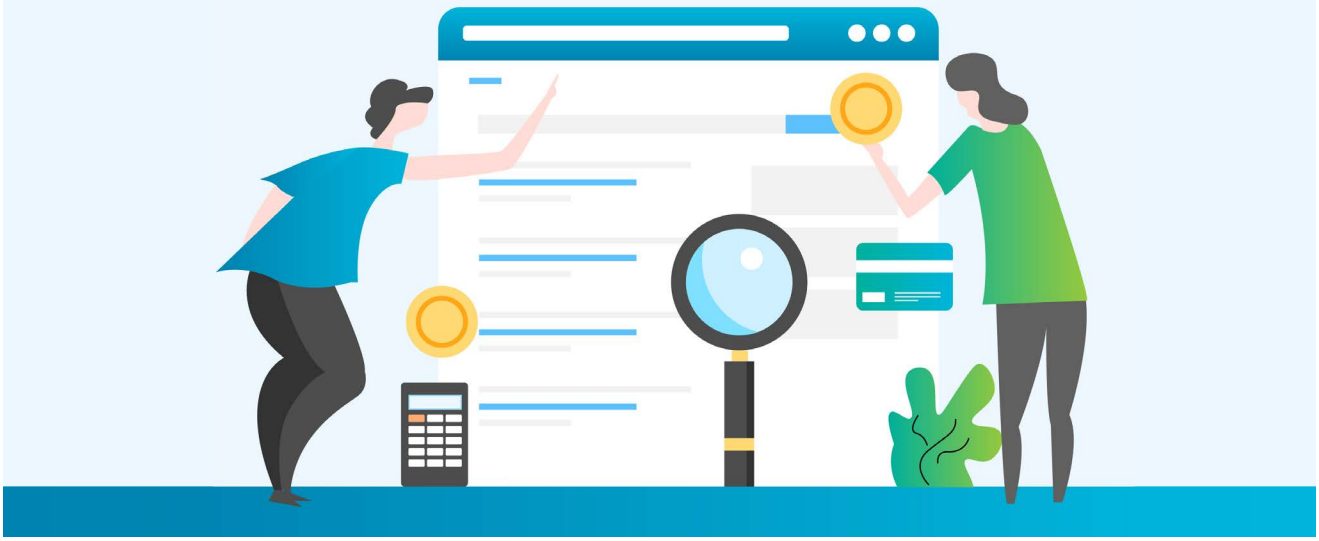
PAID SEARCH RESULTS

Not Google, not Bing, nor any other major search engine is in the business of providing organic listings.

That is to say, organic results are the means to an end, but do not directly generate revenue for them.

Without organic search results, Google's paid search results would appear less relevant (**Overture** anyone?), thus reducing eyeballs and paid clicks.

Basically, Google and Bing (and the others) are advertising engines that happen to draw users to their properties with organic listings. Organic, then, is the means to the end.



WHY DOES THIS MATTER?

It's the key point driving in:

- Their layout changes.
- The existence of search features like knowledge panels and featured snippets.
- The click-through rates (CTR) of organic results.

When Google adds a fourth paid search result to commercial-intent queries it's because of this.

When Google displays a featured snippet so you don't have to leave Google.com to get an answer to your query... it is because of this.

Regardless of what change you may see taking place it's important to keep this in mind and always question not just what it will impact today but what further changes do they imply may be on the horizon.

HOW SEARCH ENGINES WORK TODAY: THE SERIES

Alright, now that we have that baseline understanding of why Google even provides organic results let's look at the nuts-and-bolts of how they operate.

To accomplish this we're going to look at:

- Crawling and indexing
- Algorithms
- Machine learning
- User intent

This piece will focus on indexing. So let's dive in...

INDEXING

Indexing is where it all begins.

For the uninitiated, indexing essentially refers to the adding of a webpage's content into Google.

When you create a new page on your site there are a number of ways it can be indexed.

The simplest method of getting a page indexed is to do absolutely nothing.

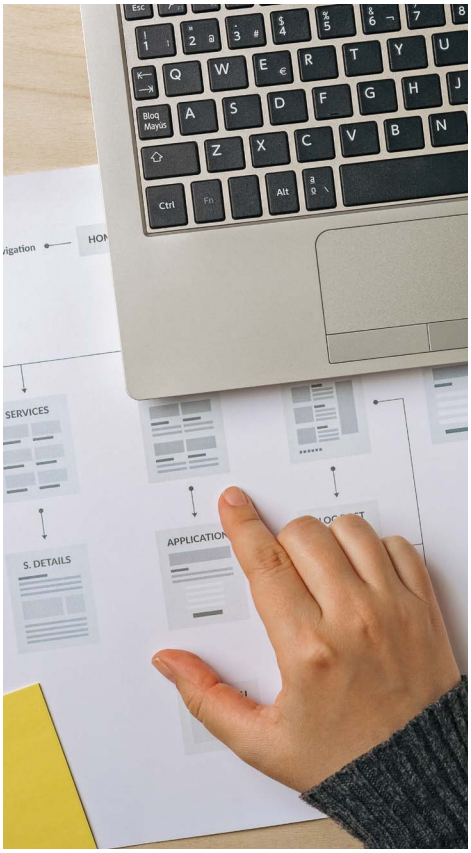
Google has crawlers following links and thus, provided your site is in the index already and that the new content is linked to from within your site, Google will eventually discover it and add it to its index. More on this later.

But what if you want Googlebot to get to your page faster?

This can be important if you have timely content or if you've made an important change to a page you need Google to know about.

One of the top reasons I use faster methods is when I've either optimized a critical page or I've adjusted the title and/or description to improve click-throughs and want to know specifically when they were picked up and displayed in the SERPs to know where the measurement of improvement starts.

In these instances there a few additional methods you can use:



1 XML SITEMAPS

There are always **XML sitemaps**.

Basically, this is a sitemap that is submitted to Google via **Search Console**.

An XML sitemap gives search engines a list of all the pages on your site, as well as additional details about it, such as when it was last modified.

Definitely recommended!

But when you need a page indexed immediately it's not particularly reliable.

2 REQUEST INDEXING

In Search Console, you can “Request Indexing”.

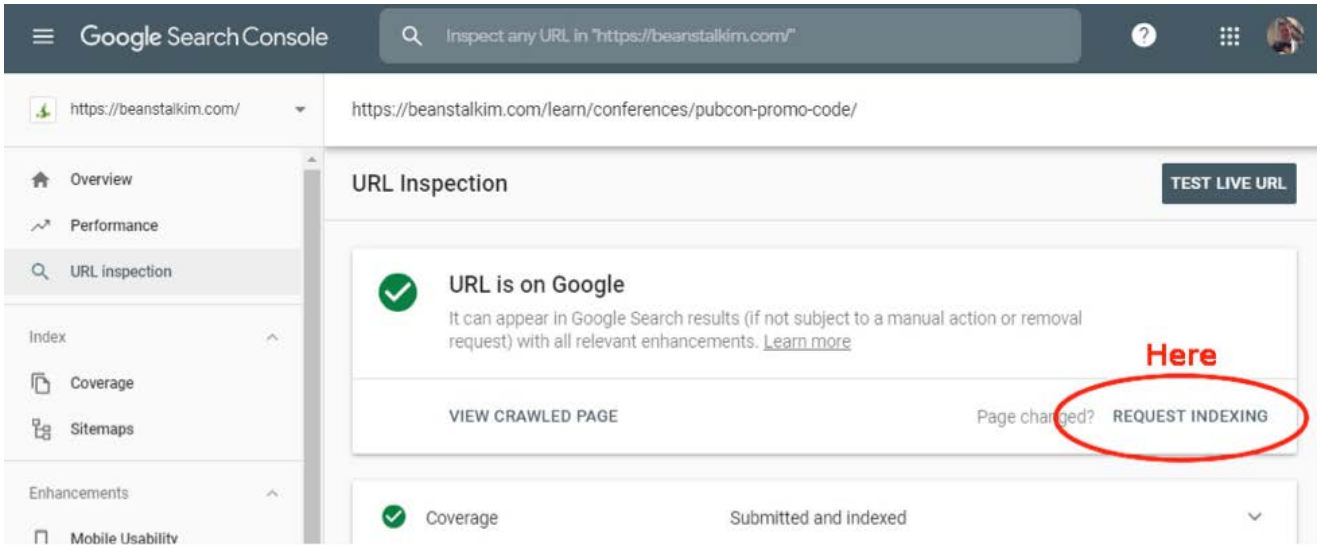
You begin by clicking on the top search field which reads by default, “Inspect and URL in domain.com”

Enter the URL you want indexed, then hit Enter.

If the page is already known to Google you will be presented with a bunch of information on it. We won’t get into that here but I recommend logging in and seeing what’s there if you haven’t already.

The important button, for our purposes here, appears whether the page has been indexed or not – meaning that it’s good for content discovery or just requesting Google to understand a recent change

You'll find the button ...



Within a few seconds to a few minutes, you can search the new content or URL in Google and find the change or new content picked up.

3 HOST YOUR CONTENT ON GOOGLE

Crawling sites to index them is a time and resource-consuming process.

One alternative is to host your content directly with them.

This can be done a few different ways but most of us (myself included) have not adopted the technologies or approaches required and Google hasn't pushed us to them.

We're seeing the ability to give Google direct access to our content via XML feeds, APIs, etc. and unplug our content from our design.

Firebase, Google's mobile app platform, gives Google direct access to the app content, bypassing any need to figure out how to crawl it.

This is the future – enabling Google to index content immediately, without effort, so it can then serve it in the format most usable based on the accessing technology.

While we aren't quite where we need to be in our technologies to stress too much about this side of things, just know it is coming.

I cannot recommend enough following Cindy Krum's **MobileMoxie blog**, where she discusses these and mobile-related subjects in great detail and with great insight.

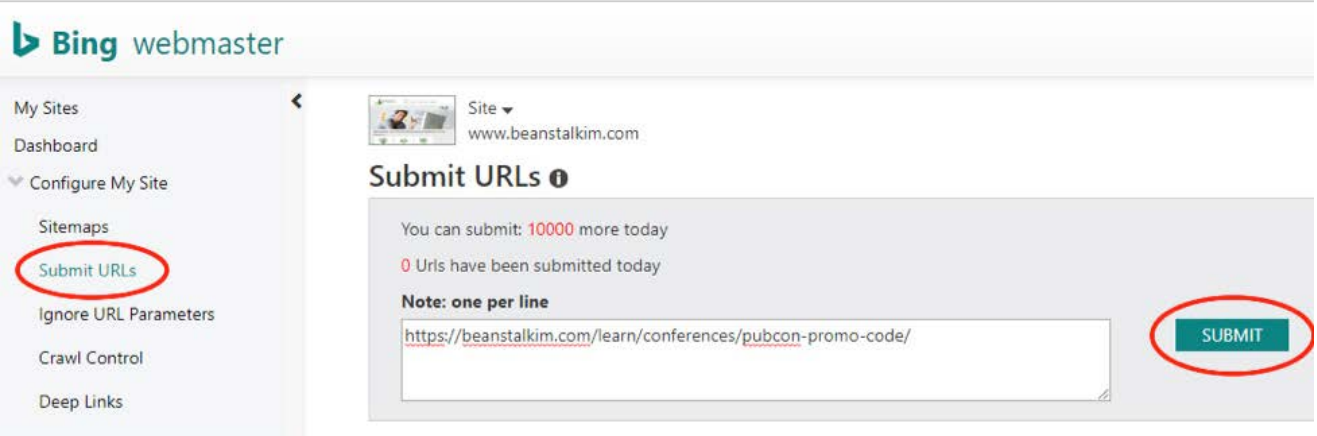
4 AND BING, TOO!

To get your content indexed and/or updates quickly by Bing, you will need a Bing Webmaster Tools account.

If you don't have one, I can't recommend it enough. The info provided within is substantial and will help you better assess problem areas and improve your rankings on Bing, Google and anywhere else – and probably provide a better user experience as well.

But for getting your content indexed you simply need click: Configure My Site > Submit URLs

From there you enter the URL(s) you want indexes and click "Submit".



So – that's almost everything that you need to know about indexing and how search engines do it (with an eye towards where things are going).

CRAWL BUDGET

We can't really talk about indexing without talking about crawl budget.

Basically, crawl budget is a term used to describe the amount of resources that Google will expend crawling a website.

The budget assigned is based on a combination of factors, the two central ones being:

• How fast your server is (i.e., how much can Google crawl without degrading your user experience).

• How important your site is.

If you run a major news site with constantly updating content that search engine users will want to be aware of your site will get crawled frequently (dare I say … constantly).

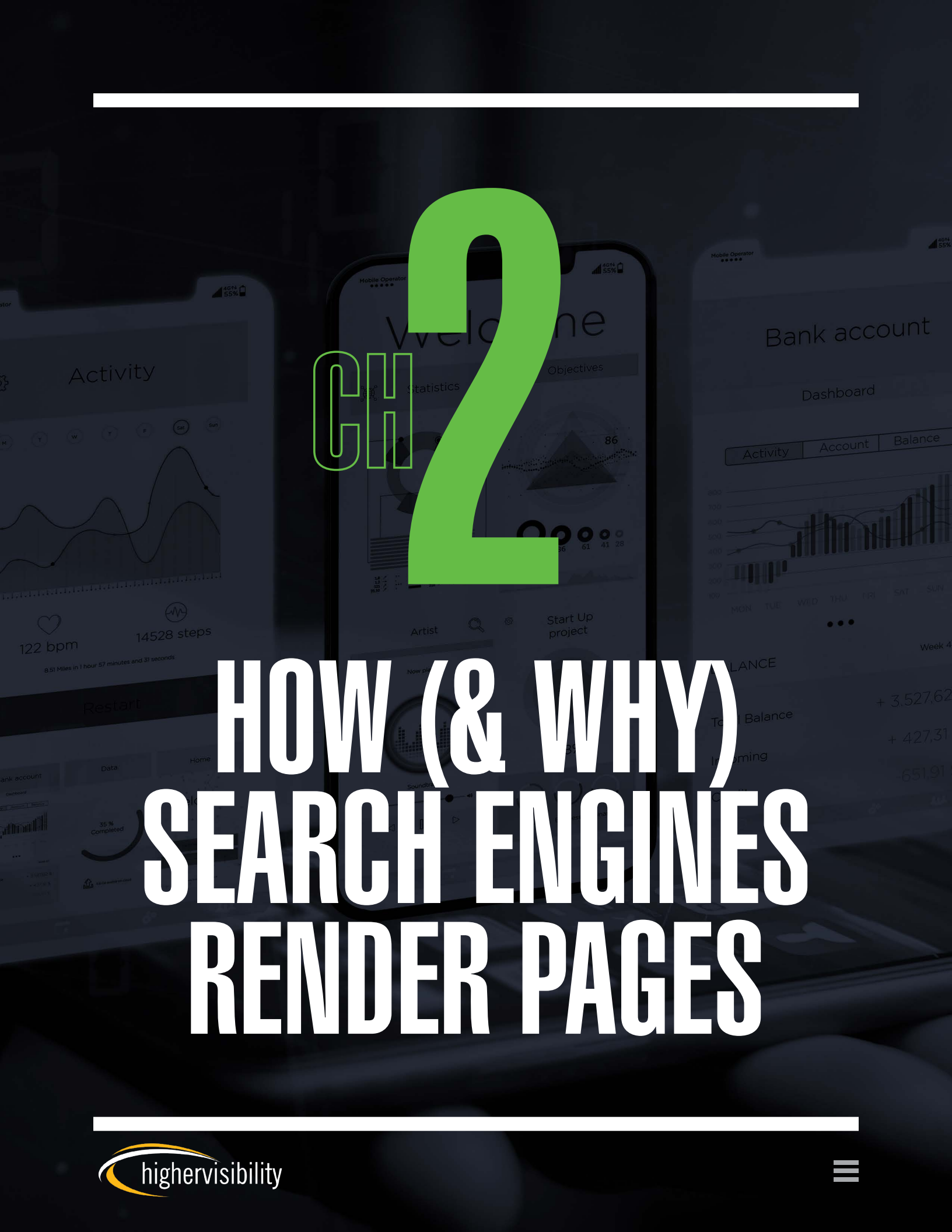
If you run a small barbershop, have a couple of dozen links, and rightfully are not deemed important in this context (you may be an important barber in the area but you're not important when it comes to crawl budget) then the budget will be low.

You can read more about crawl budgets and how they're determined in Google's explanation **here.**

TO BE CONTINUED...

So now we understand crawling and the search indexes.

Part 2 will look in more detail at where rendering comes in...


CH 2
HOW (& WHY) SEARCH ENGINES RENDER PAGES
Activity
122 bpm
14528 steps
Restart
Bank account
Dashboard
Activity
Account
Balance
Objectives
Statistics
Artist
Start Up project
highervisibility

There is an interesting twist in how we think about indexing – and that is rendering.

When we think about ranking pages we generally think about indexing. This is to say, we generally think about the point in time when a search engine has:

- Discovered a page through sitemaps or crawling and has proceeded to then visit the page for indexing.
- Gathered all the content on the page.
- Started ranking the page for queries.

Arguably, this is the most important stage in the process given that this is the trigger for rankings, but it's not the final stage of the discovery process and I would suggest that its weight will decline in time while the final stage – rendering – gains traction.

WHAT IS THE DIFFERENCE BETWEEN INDEXING & RENDERING?

Essentially, the difference between indexing and rendering can be illustrated with these two images:

This is basically the same content, viewed as it would be during indexing (HTML) and rendering (Chrome).

WHY DOES THIS MATTER?

Now, you may be asking yourself why this matters.

If you are, then I'll assume you don't have a JavaScript site but even if that's true, it's more important than you might think. The fact that search engines rendered pages prior to the recent push into JavaScript use for websites is a good confirmation.

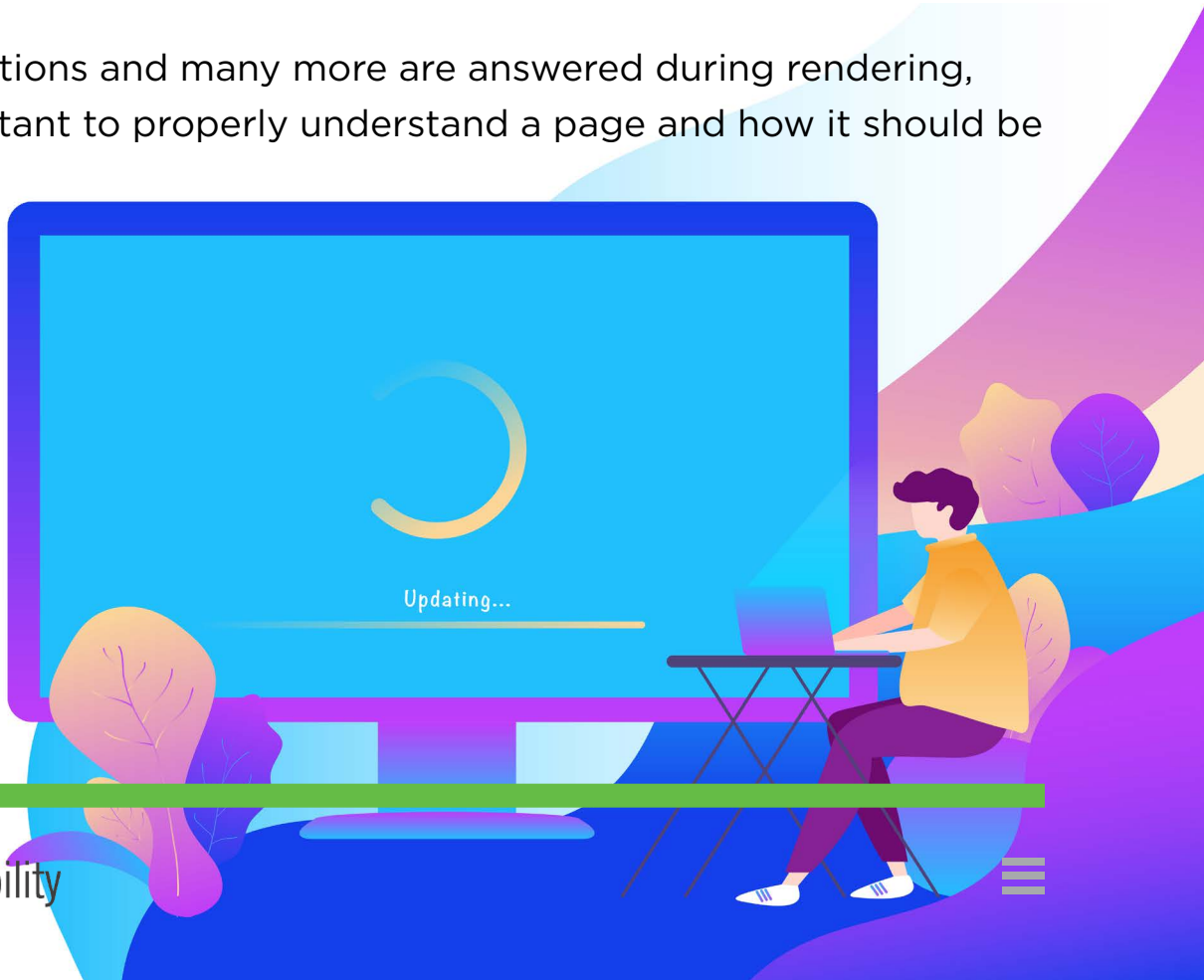
Essentially the reason that it matters is that rendering provides the truth.

With the code, a search engine can understand what a page is about and roughly what's going on.

With rendering, they can understand the user experience and far more about what content should take priority.

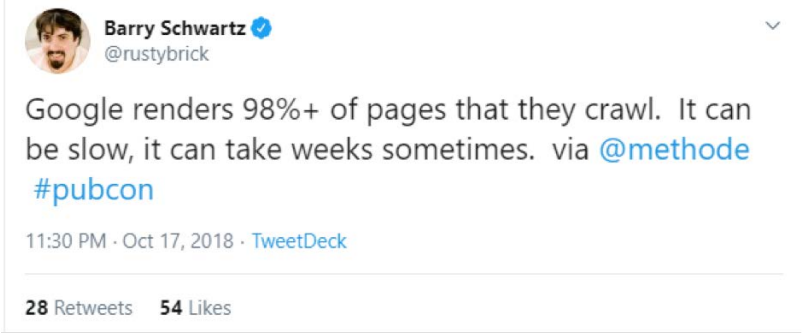
- Is content hidden behind a click?
- Does an ad fill the page?
- Is content that appears towards the bottom of the code, actually displayed towards the top or in the navigation?
- Is a page slow to load?

All these questions and many more are answered during rendering, and are important to properly understand a page and how it should be ranked.



WHEN DOES RENDERING OCCUR?

Rendering occurs after indexing. How long after is not set in stone, but according to Gary Illyes from Google **it can take several weeks.**



When I asked John Mueller of Google if this timeline was still accurate today the response was:

So, it's something that they're actively working on.

Bing operates differently of course, but according to their Web Ranking & Quality Project Manager, Frédéric Dubut, **the timeline is roughly the same.**

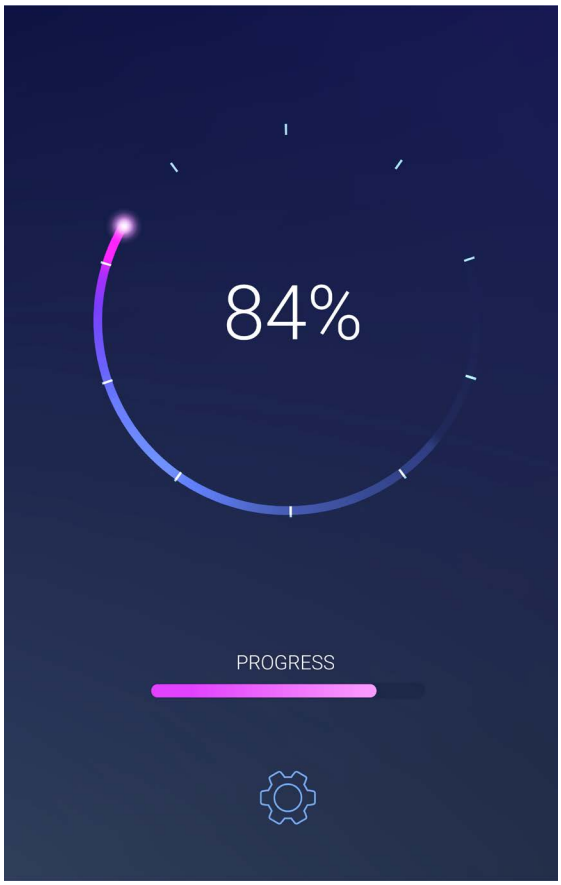
So, the short answer is "after indexing" and the timeline is variable, essentially meaning that the search engines will understand the content and context of a page prior to gaining a full understanding of how it is to be prioritized.

This is not to say that they are completely ignorant until rendering.

There are some solid rules and understandings that the engines have all gained over the years that allow them to make quick assumptions about:

- What elements do.
- Where they're positioned.
- How important they are meant to be to the user.

But it isn't until the pages are rendered that the engines will know their assumptions are correct and that they can fully understand a page and its form.



THE PROBLEM WITH RENDERING

In essence, the search engines send a crawler to the site that will render the page as a browser would.

Based on its popularity, we will use Google as an example here.

Googlebot has a Web Rendering Service (WRS) component. Thankfully, this component was **updated in May of 2019.**

Until then, the Web Rendering Service was using Chrome version 41. While this was great for compatibility it was a nightmare for sites that relied on modern features like those in modern JavaScript.

In May 2019, the Web Rendering Service was upgraded to evergreen, meaning that it uses the most current version of Chrome for rendering (within a couple weeks at any rate).

Essentially, now when your page is rendered by Googlebot, it's rendered more-or-less how you would see it in your browser.

Great right? Now the only testing you need to do is open a browser and if it works there it's ready for Google, right? Right?

You can probably guess the answer. Wrong.

And Bing isn't much better (though they do seem to be a bit better at rendering which is interesting).

If you have a basic site with predictable HTML and little-to-no dynamic content, then there really isn't anything you need to worry about and there probably wasn't with the old Web Rendering Service setup either.

But for those with dynamic content served via JavaScript, there is a very big caveat and it's rooted in this gap.

Namely, until the page is rendered, the engine doesn't know what's

on it. Unlike a site with a simple HTML output where the engine might be missing a bit of the context but has the content, with a site built on something like JavaScript that relies on the rendering, the engine will not know what content is on the page until the Web Rendering Service has done its job.

Suddenly those "weeks" are pretty impactful. This is also why the engines are working to reduce the latency.

Until they do, JavaScript developers will need to rely on pre-rendering (creating a static version of each page for the engines) which is not at all ideal.

WHAT DOES A WEB RENDERING SERVICE DO?

I wanted to quickly answer a question that I found myself not quite wrapping my brain around until I realized I was thinking about it entirely wrong. You are welcome to laugh at me for the obviousness of the hiccup in my brain.

First, let's consider where a Web Rendering Services gets its instructions and how.

HERE'S BASICALLY THE LIFE-CYCLE OF RENDERING:

- A page is discovered via sitemap, crawler, etc.
- The page is added to the list of pages to be crawled on a site when the crawl budget is available.
- The page content is crawled and indexed.
- The page is added to the list of pages to be rendered on a site when the rendering budget is available.
- The page is rendered.

So, a critical and unspoken element of the process is the rendering queue. Googlebot may get to a page weeks before rendering it and until then some content (JavaScript sites) or context (all sites) may be missing.

When a page hits the top of the queue for rendering, the engine will send what is referred to as a headless browser to it.

This is the step I had difficulty with. A headless browser is a browser without a graphical user interface.

For some reason, I had a difficult time wrapping my brain around how that worked. Like, how is Google to know what's there if it's not graphically displayed?

The obvious answer is of course:

“The bot doesn’t have eyes either so … um … yeah.”

Over that mental hiccup, I came to terms with it as a “browser light” that renders the page for the search engine to now understand what appears where and how on a page – even though they don’t have eyes to see it.

When all goes well, the rendered version will appear the same to Googlebot as it does to graphical browsers and if it doesn’t then it’s likely because the page relies on an unsupported feature like a user permission request.

ALL IN ALL. . .

I suspect that in the coming months, we will see the latency between indexing and rendering shrink dramatically, especially on sites that rely on it.

This won’t have a dramatic impact on most sites but for those that need to be rendered to be understood … the world may open up.

Though more likely, a new set of problems and hiccups will unfold.

Because from my experience, we can count on the indexing skills of the engines, but the rendering side still has a long way to go in bridging the gap between what the search engines see and what a user’s browser does.

CH 3

HOW SEARCH ENGINE ALGORITHMS WORK:

EVERYTHING YOU NEED TO KNOW

SEJ Search Engine Journal

Often I find myself focusing on specific strategies to perform specific functions.

How do I write compelling copy to **rank on voice search?**

What **structured data produces easy wins?**

Things like that.

These important questions are often covered here on Search Engine Journal in very useful articles.

But it's important to not just understand what tactics might be working to help you rank. You need to understand how it works.

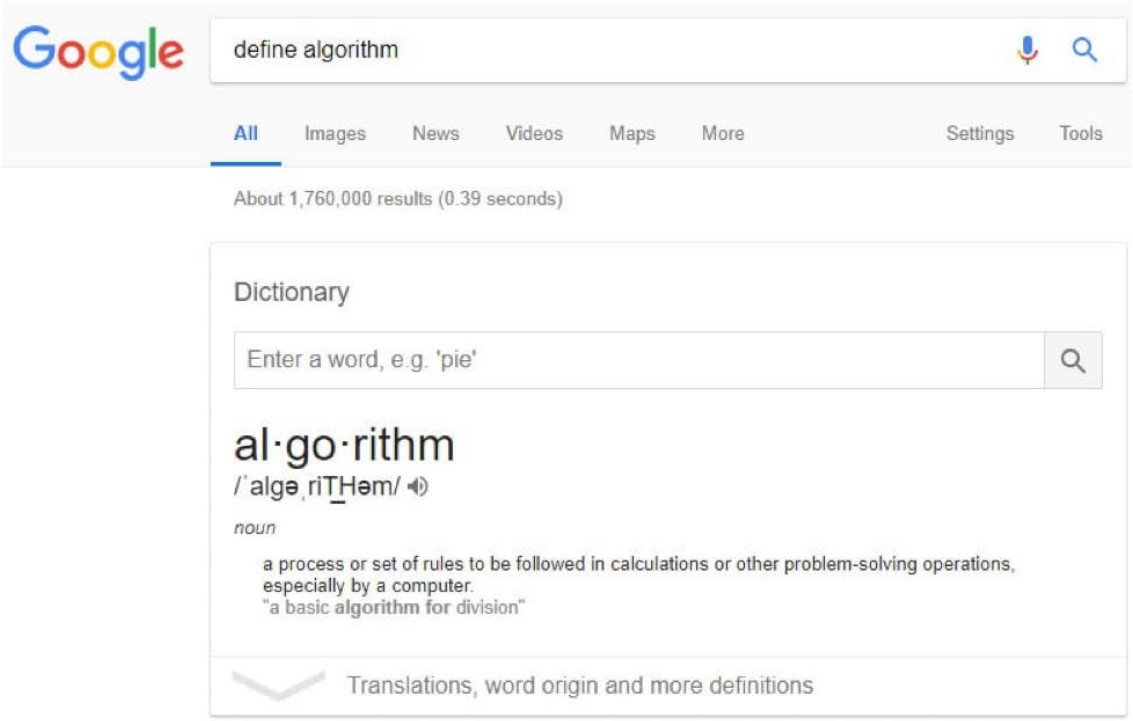
Understanding the structure that the strategy is functioning in is paramount to understanding not just why that strategy is working, but how and what it's trying to accomplish.

Previously, we discussed **how search engines crawl and index information.**

This chapter will explore the basics of how search algorithms work.

WHAT IS AN ALGORITHM? A RECIPE

If you ask Google what an algorithm is, you'll discover that the engine itself (and pretty much everyone else) defines it as "a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer."



If you take anything from this definition, it's critical to understand what it is not in our context here.

AN ALGORITHM IS *NOT* A FORMULA.

To wrap our heads around the difference, why it's important, and what each does, let's consider for a moment the meal I might place on my dinner plate tonight.

We'll go with a favorite of mine:

- Roast beef
- Horseradish
- Yorkshire pudding
- Green beans
- Mashed potatoes
- Gravy

(That's right, we Canadians eat more than poutine and maple syrup, though both are awesome though probably not together.)

The roast beef needs to be seasoned and cooked perfectly. The seasoning combined with the roast would be an example of a formula – how much of each thing is necessary to produce a product.

A second formula used would be the amount of time and at what temperature the roast should be cooked, given its weight. The same would occur for each item on the list.

At a very basic level, we would have 12 formulas (6 items x 2 – one for measurements and the other for cooking time and duration based on volume) making an algorithm set with the goal of creating one of Dave's favorite meals.

We aren't even including the various formulas and algorithms required to produce the ingredients themselves, such as raising a cow or growing potatoes.

Let's add one more formula though – a formula to consider the amount of different foods I would want on my plate.

So, we now have an algorithm to accomplish this very important task. Fantastic! Now we just need to personalize that algorithm so that the rest of my family also enjoys their meal.

We need to consider that each person is different and will want different amounts of each ingredient and may want different seasonings.

So, we add a formula for each person. Alright.

AN ALGORITHM OF ALGORITHMS

What the heck does a search algorithm and a dinner table have in common?

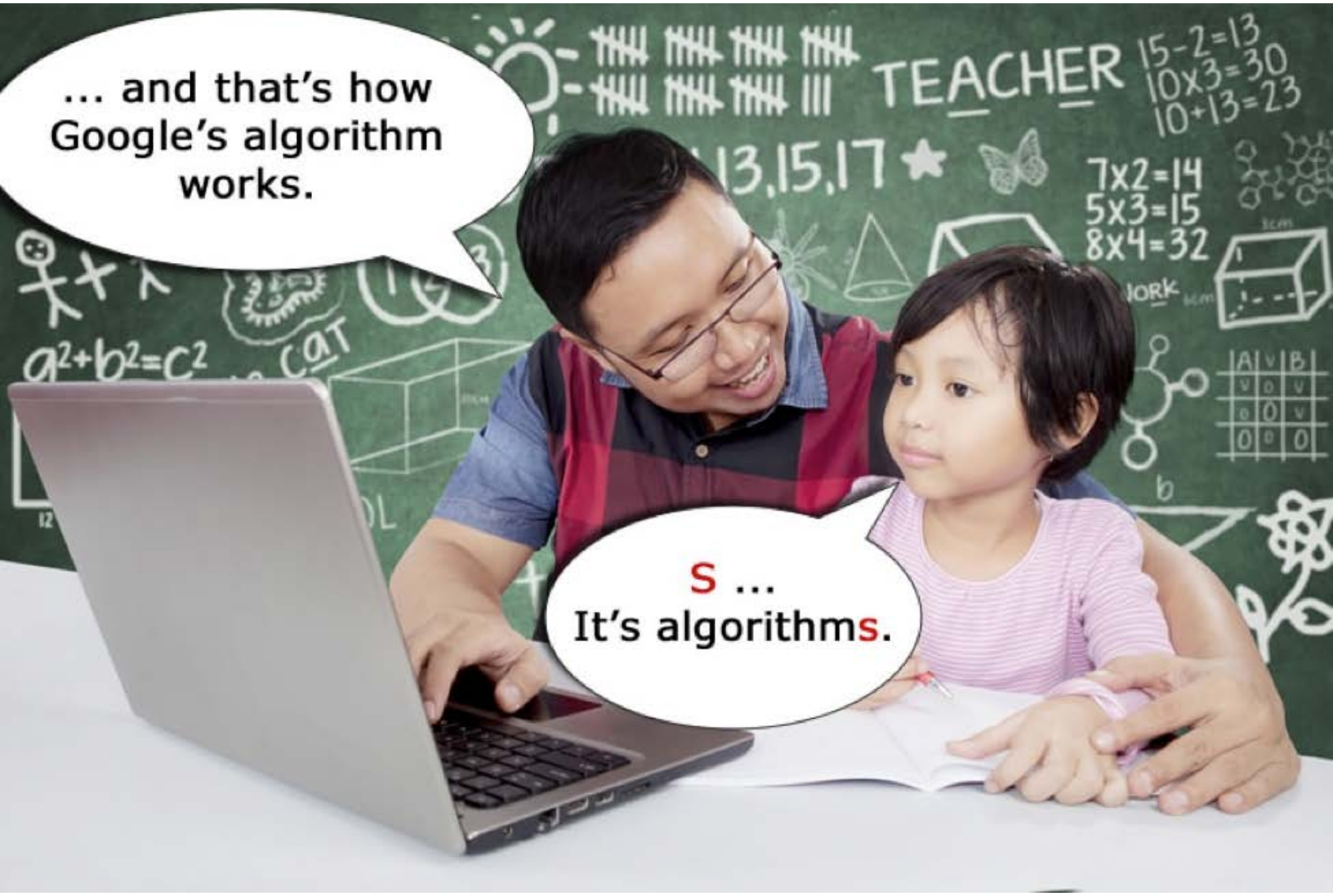
A lot more than you think.

Let's look at just a few of the core characteristics of a website for comparison. ("Few" meaning nowhere near everything. Like not even close.)

- URLs
- Content
- Internal links
- External links
- Images
- Speed

As we witnessed with our dinner algorithm, each of these areas is divided further using different formulas and, in fact, different sub-algorithms.

It might be better if we think of it not as an algorithm, but as algorithms.



It's also important to keep in mind that, while there are many algorithms and countless formulas at play, there is still an algorithm.

Its job is to determine how these others are weighted to produce the final results we see on the SERP.

So, it is perfectly legitimate to recognize that there is some type of algorithm at the top – the one algorithm to rule them all, so to speak – but always recognize that there are countless other algorithms and generally they're the algorithms we think about when we're considering how they impact search results.

Now, back to our analogy.

We have a plethora of different characteristics of a website being rated just as we have a number of food elements to end up on our dinner plate.

To produce the desired result, we have to have a large number of formulas and sub-algorithms to create each element on the plate and master algorithm to determine the quantity and placement of each element.

Sound familiar?

When we're thinking of "Google's algorithm" what we're actually referring to is a massive collection of algorithms and formulas, each set to fulfill one specific function and gathered together by a lead or, dare I say, "core" algorithm to place the results.

So, we have:

• Algorithms like **Panda** to assist Google in judging, filtering, penalizing and rewarding content based on specific characteristics and that algorithm likely included a myriad of other algorithms within in.

• The **Penguin** algorithm to judge links and address spam there. But this algorithm certainly requires data from other pre-existing algorithms that are responsible for valuing links and likely some new algorithms tasked with understanding common link spam characteristics so the larger Penguin algorithm could do its job.

• Task-specific algorithms.

• Organizing algorithms.

• Algorithms responsible for collecting all the data and putting it into a context that produces the desired result, a SERP that users will find useful.

So there we have it. That's how search algorithms work at their core.

WHY SEARCH ALGORITHMS USE ENTITIES

One of the areas of search that's getting some decent attention lately, though which is under-emphasized, is the idea of **entities.**

For context, an entity is defined by Google as:

"A thing or concept that is singular, unique, well-defined and distinguishable."

So, in our dinner analogy, there's me. I'm an entity.

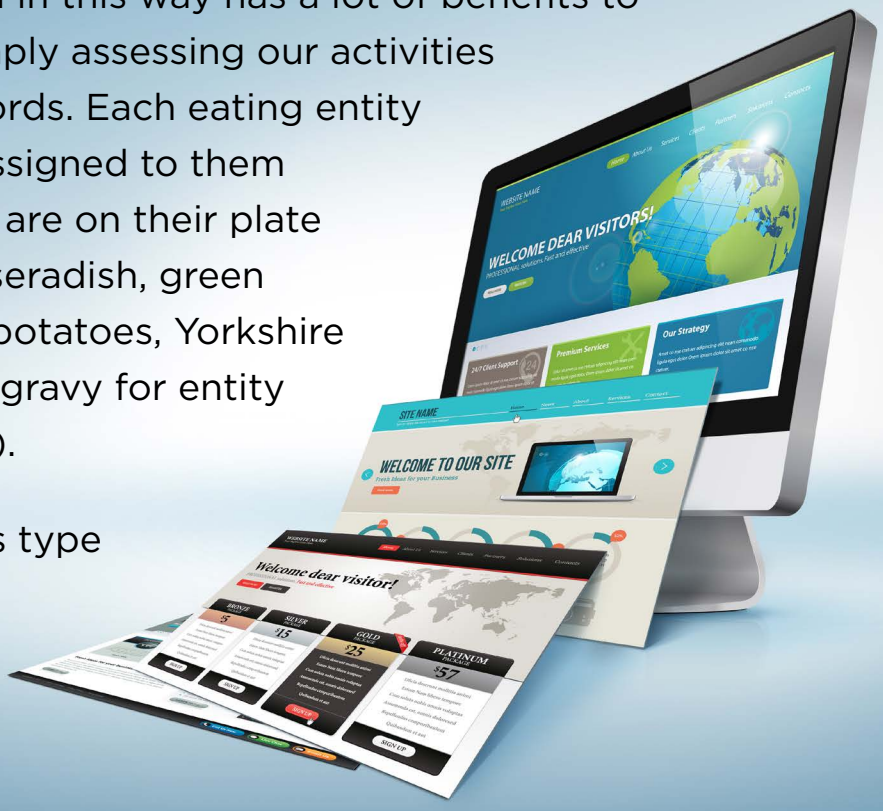
Each member of my family is also their own entity. In fact, my family unit is an entity unto itself.

By that token, the roast and each ingredient that goes into it are also their own entities. So is the Yorkshire pudding and so is the flour that went into making it.

Google sees the world as a collection of entities. Here's why:

At my dinner table, I have four individual entities that would have the state "eating" and a host of entities being consumed.

Classifying us all in this way has a lot of benefits to Google over simply assessing our activities as a series of words. Each eating entity can now have assigned to them the entities that are on their plate (roast beef, horseradish, green beans, mashed potatoes, Yorkshire pudding but no gravy for entity xyz1234567890).



Google uses this type of classification to judge a website.

Think of each entity sitting at the table as a page. The global entity that represents us all (let's call this entity "Davies") would be about "roast beef dinner," but each individual entity representing an individual (or page in our analogy) is different.

In this way, Google can easily classify and judge the interconnectedness of websites and the world at large.

Basically, search engines aren't responsible to just judge one website – they must rank them all. The entity "Davies" is seen to be about "roast beef dinner" but the entity next door (let's call this entity "Robinsons") is about "stir fry."

Now if an outside entity known as "Moocher" wanted to determine where to eat, the options can be ranked to Moocher based on their preferences or query.

Where (in my opinion) the real value in entities lies is in what happens the day after. We have some leftovers. By processing the entity "roast beef" with a different formula and adding the entities bread, cheese, and onions, we have:

HOW SEARCH ALGORITHMS USE ENTITIES

OK, it may not seem obvious how important this is in understanding search algorithms and how entities work in this way. While understanding how Google seeing what a website is about as a whole has an obvious value, you may be asking why it's relevant for Google to understand that my roast beef and beef dip are related and in fact – are drawn from the same core entity.

Let's consider instead Google understanding that a webpage is about roast beef. Let's also consider that another page links to it and that page is about beef dip.

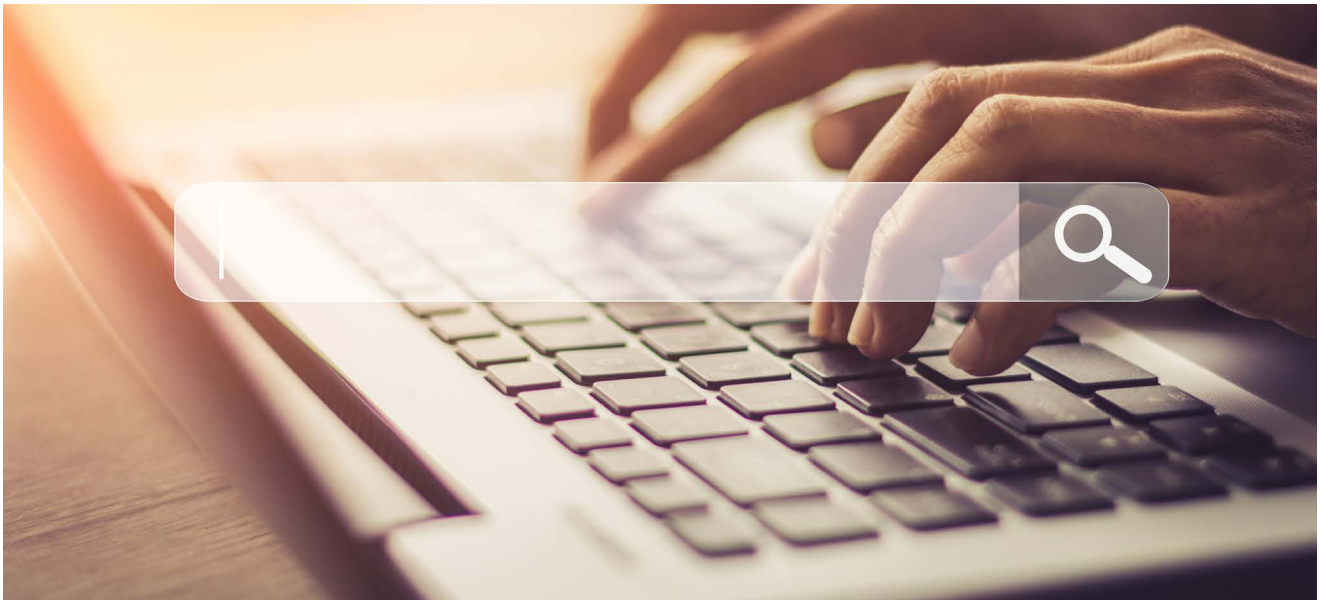
In this scenario, it's incredibly important that Google knows that roast beef and beef dip are drawn from the same core entity. They can assign relevance to this link based on the connectedness of these entities.

Before the idea of entities entered search, engines were left to assign relevance based on word proximity, density, and other easily misinterpreted and manipulated elements.

Entities are far more difficult manipulate.

Either a page is about an entity or it's not.

Through crawling the web and mapping common ways that entities relate, search engines can predict which relationships should carry the greatest weight.



SO, HOW DO SEARCH ALGORITHMS WORK?

Alright, we've covered a lot of ground and you're probably getting hungry. You want some takeaways.

CONTEXT MATTERS

It's important to understand how algorithms function to apply context to what you're experiencing/reading.

When you hear of an **algorithm update**, it's important to know that what is being updated is likely a small piece of a very large puzzle.

Knowing this assists in interpreting which aspects of a site or the world are being adjusted in an update and how that adjustment fits into the large objective of the engine.

ENTITIES ARE SUPER IMPORTANT

Further, it's critical moving forward to understand that entities:

- Play a massive role in search algorithms today.
- Have their own algorithms.
- Will play an ever-increasing role over time.

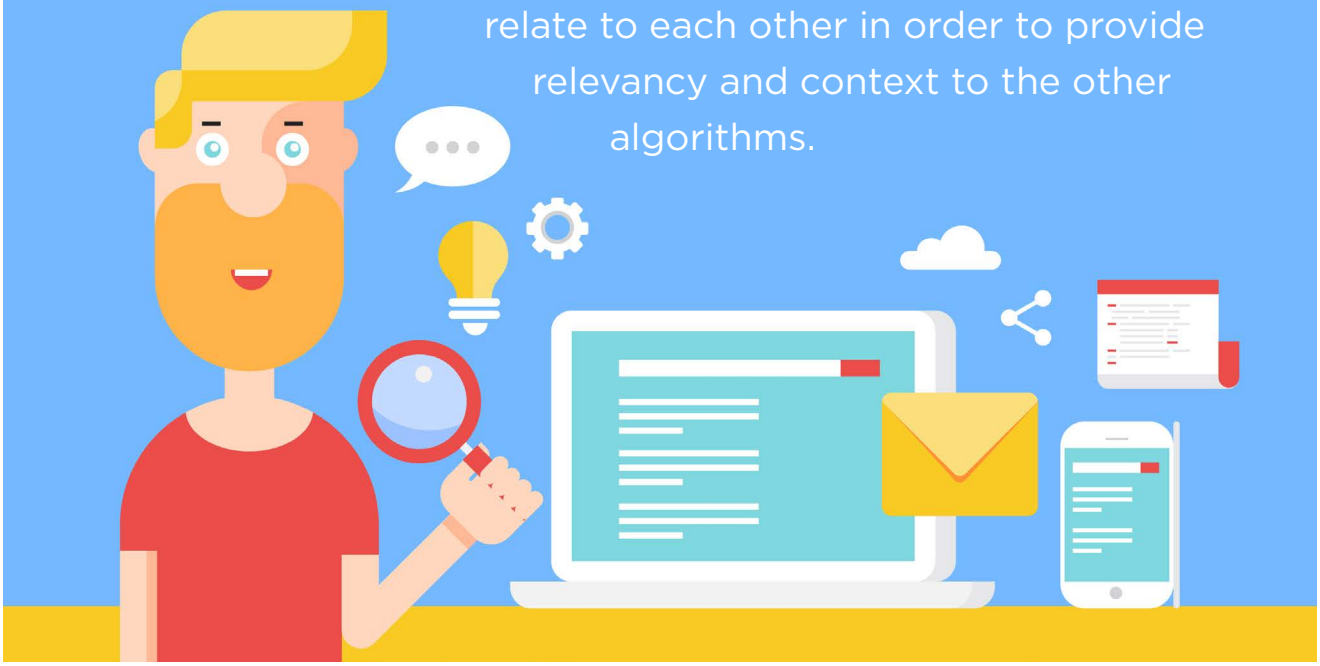
Knowing this will help you understand not just what content is valuable (how close are those entities you're writing about?) but also which links are likely to be judged more favorably. And that's just to name a couple of advantages.

IT'S ALL ABOUT USER INTENT

Search algorithms work as a large collection of other algorithms and formulas, each with its own purpose and task, to produce results a user will be satisfied with.

In fact, there are algorithms in place to monitor just this aspect of the results and make adjustments where ranking pages are deemed not to satisfy user intent based on how users interact with it.

Included in this are algorithms designed specifically to understand entities and how entities relate to each other in order to provide relevancy and context to the other algorithms.



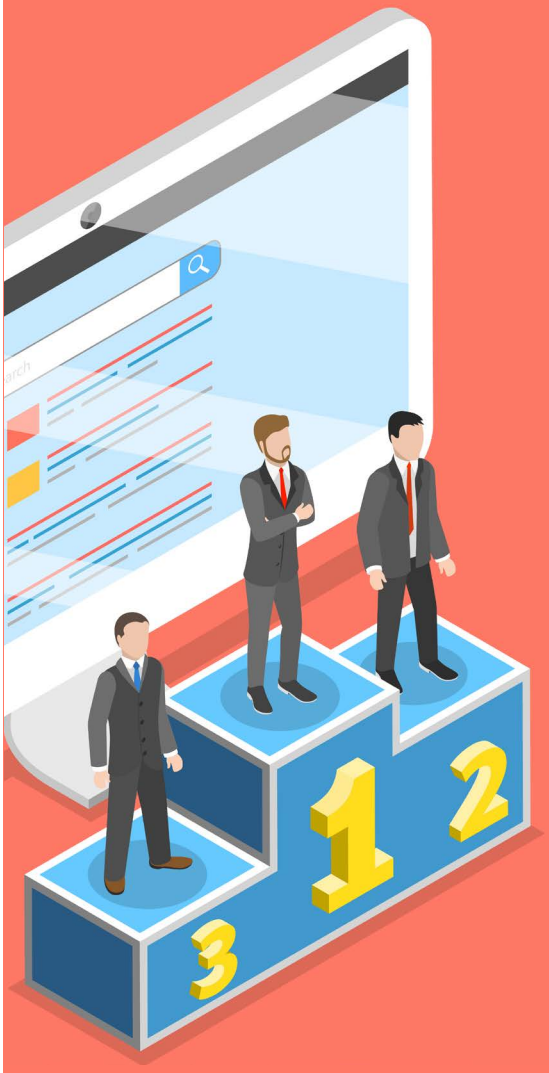
CH 4

HOW SEARCH ENGINES RANK PAGES

As SEO professionals, we generally focus on the question, "How do I rank my page?"

An equally, if not more important question we should be asking is, "How do search engines rank pages?"

WHY SEARCH ENGINES RANK WEB PAGES



Before we dive into how search engines rank web pages, let's stop for a moment and think about why they rank them.

After all, it would be cheaper and easier for them to simply display pages randomly, by word count, by freshness, or any of a variety of easy sorting systems.

The reason they don't do that is obvious. You wouldn't use it.

So when we ask the question about rankings, what we need to always keep in mind is that the user we are trying to satisfy is not ours, they belong to the engine and the engines are loaning them to us.

If we misuse that user, they may not return to the engine and thus the engine can't have that as their ad revenue will decline.

I like to think of the scenario like some of the resource pages on our own site.

If we recommend a tool or service, it is based on our experience with them and we believe they will serve our visitors as well. If we hear they do not, then we will remove them from our site.

That's what the engines are doing.

But how?

DISCLAIMER

I do not have eavesdropping devices at Google or Bing. Google has one sitting on my desk and another I carry around with me when I'm not at it, but for some reason, the message pickup doesn't work the other way.

I state this to make clear that the following outline is based on about 20 years of watching the search engines evolve, reading patents (or more often – **Bill Slawski**'s analysis of patents), and starting each day for many years by reviewing the goings-on in the industry from SERP layout changes to acquisitions to **algorithm updates.**

Take what I am saying as an educated breakdown that's hopefully about 90% right.

If you're wondering why I think 90% – I learned from Bing's **Frédéric Dubut** that **90%** is a great number to use when guesstimating.

IT'S ONLY A SIMPLE 5 STEPS – EASY

There are five steps to the complete process of ranking a page.

I am not including the technical challenges like load-balancing and I'm not talking about each various signal calculation.

I'm just talking about the core process that every query needs to go through to start its life as an information request and end it as a set of 10 blue links buried beneath a sea of ads.

Understand this process, understand who it is designed to serve, and you will be on your way to thinking properly about how to rank your pages to their users.

I also feel it's necessary to note that the words used for these steps are mine and not some type of official name.

Feel free to use them but don't expect any one of the engines to use the same terminology.

STEP 1

CLASSIFY

The first step in the process is to classify the query coming in. The classification of the query gives the engine the information it needs to perform all of the following steps.

Before complex classification could take place (read: back when the engines relied on keywords instead of entities) the engines basically had to apply the same signals to all queries. As we will explore further below, this is no longer the case.

It's in this first stage that the engine will apply such labels (again, not a technical term but an easy way to think about it) to a query such as:

- YMYL
- Local
- Unseen
- Adult
- Question

I have no idea how many different classifications there are but the first step the engine would need to make is to determine which apply to any given query.

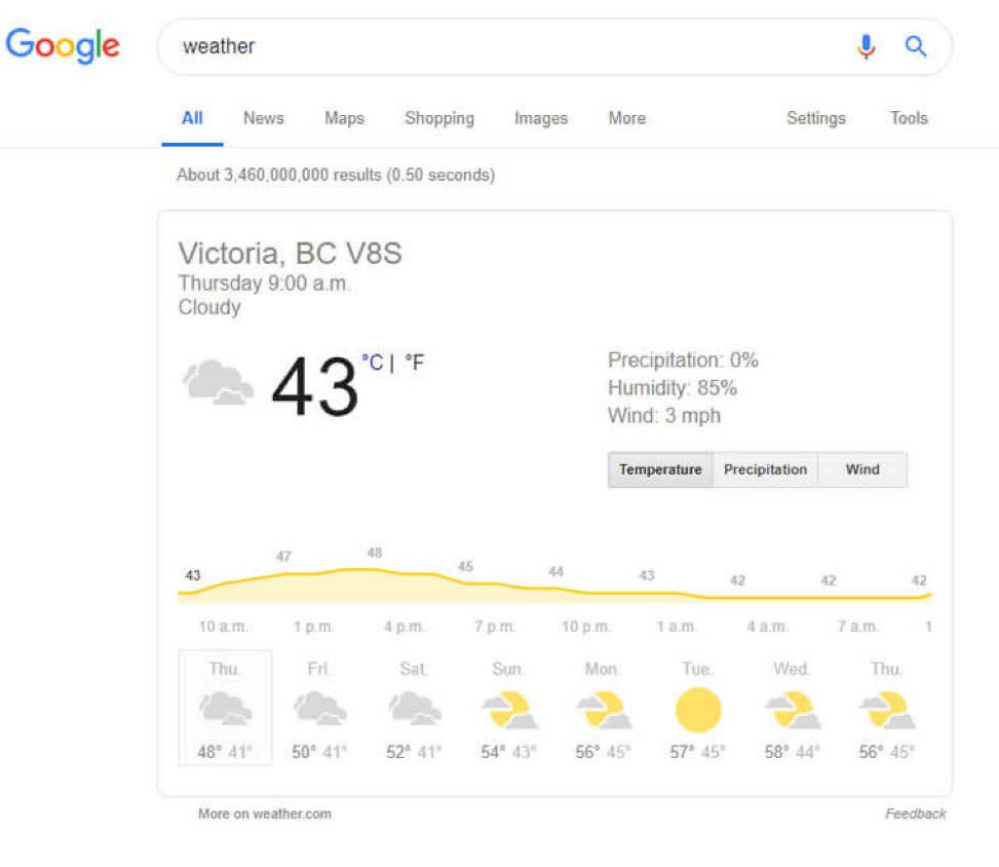
STEP 2

CONTEXT

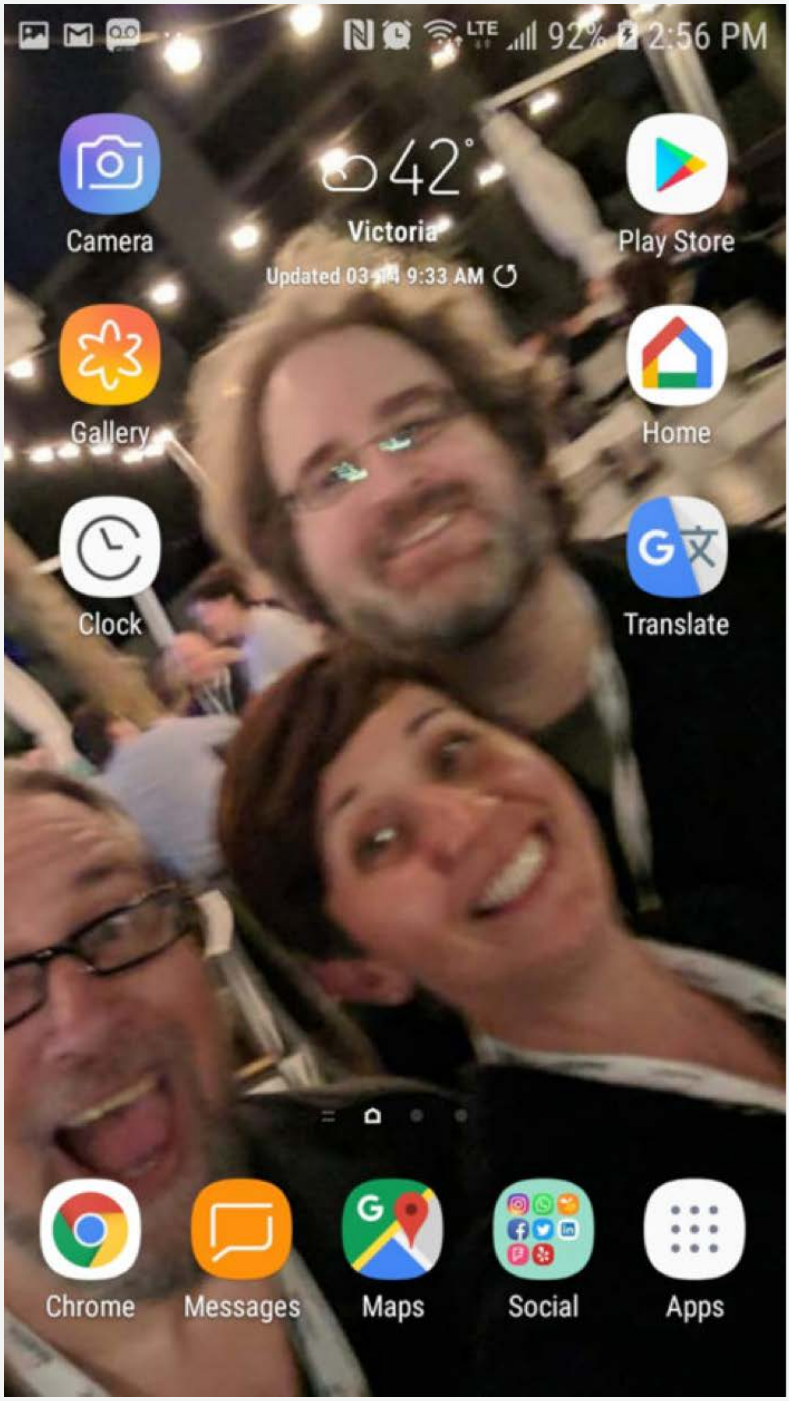
The second step in the ranking process is to assign context.

Where possible, the engine needs to take into account any relevant information they have on the user entering the query.

We see this regularly for queries, even those we don't ask. We see them here:



And we see them here:



The latter, of course, being an example of where I didn't specifically enter the query.

Essentially, the second stage in the process is for the engine to determine what environmental and historical factors come into play.

They know the category of the query, here they apply, determine or pull the data related to elements deemed relevant for that query category and type.

Some examples of environmental and historical information that would be considered are:

- Location
- Time
- Whether the query is a question
- The device being used for the query
- The format being used for the query
- Whether the query relates to previous queries
- Whether they have seen that query before

STEP 3

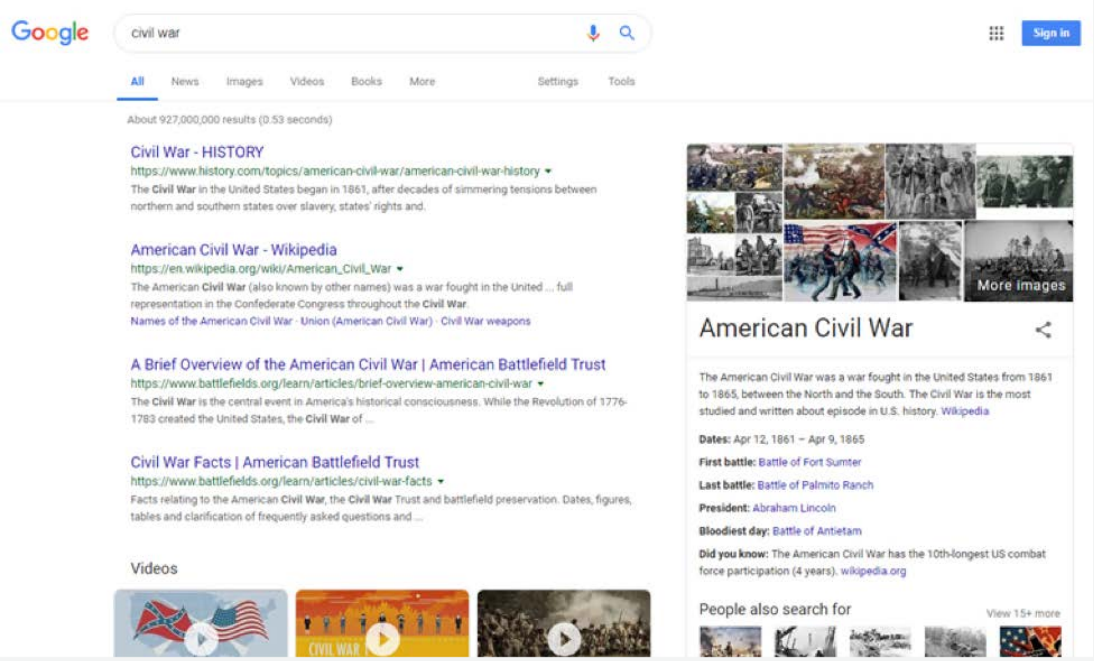
WEIGHTS

Before we dive in let me ask you, how sick are you of hearing about **RankBrain**?

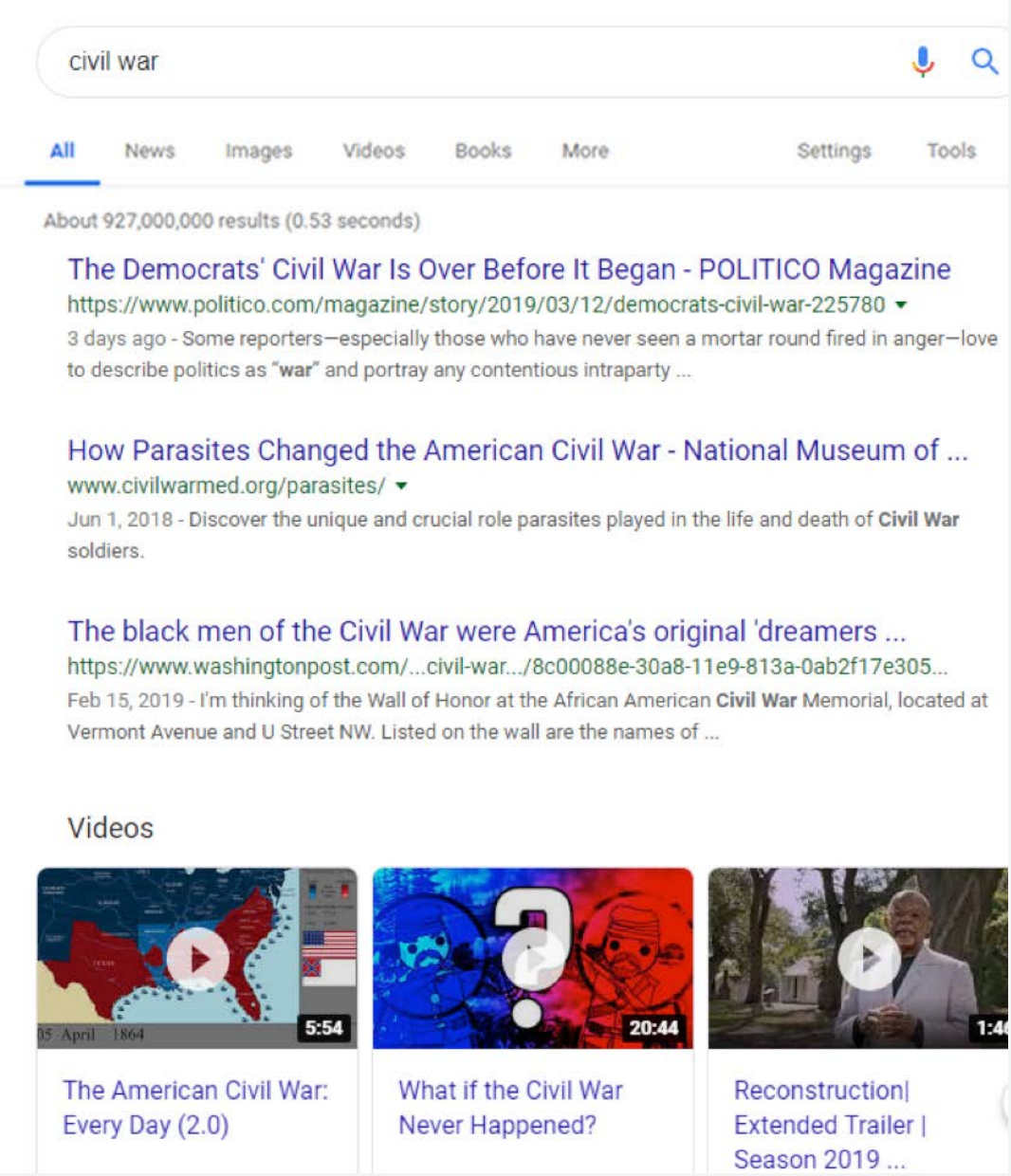
Well, buckle up because we're about to bring it up again but only as an example of this third step.

Before an engine can determine what pages should rank they first need to determine which signals are most important.

For a query like [civil war] we get a result that looks like:



Solid result. But what happens if freshness had played a strong role? We'd end up with a result more like:



But we can't rule out **freshness**. Had the query been [best shows on netflix], I'd care less about authority and more about how recently it was published.

I hardly want a heavily linked piece from 2008 outlining the best DVDs to order on their service.

So, with the query type in hand as well as the context elements pulled the engine can now rely on their understanding of which of their signals applies and with which weightings for the given combinations.

Some of this can certainly be accomplished manually by the many talented engineers and computer scientists employed and part of it will be handled by systems like RankBrain which is (for the 100th time) a machine learning algorithm designed to adjust the signal weights for previously unseen queries but later introduced into Google's algorithms as a whole.

With the statement that roughly 90% of its ranking algorithms rely on machine learning, it can reasonably be assumed that Bing has similar systems.

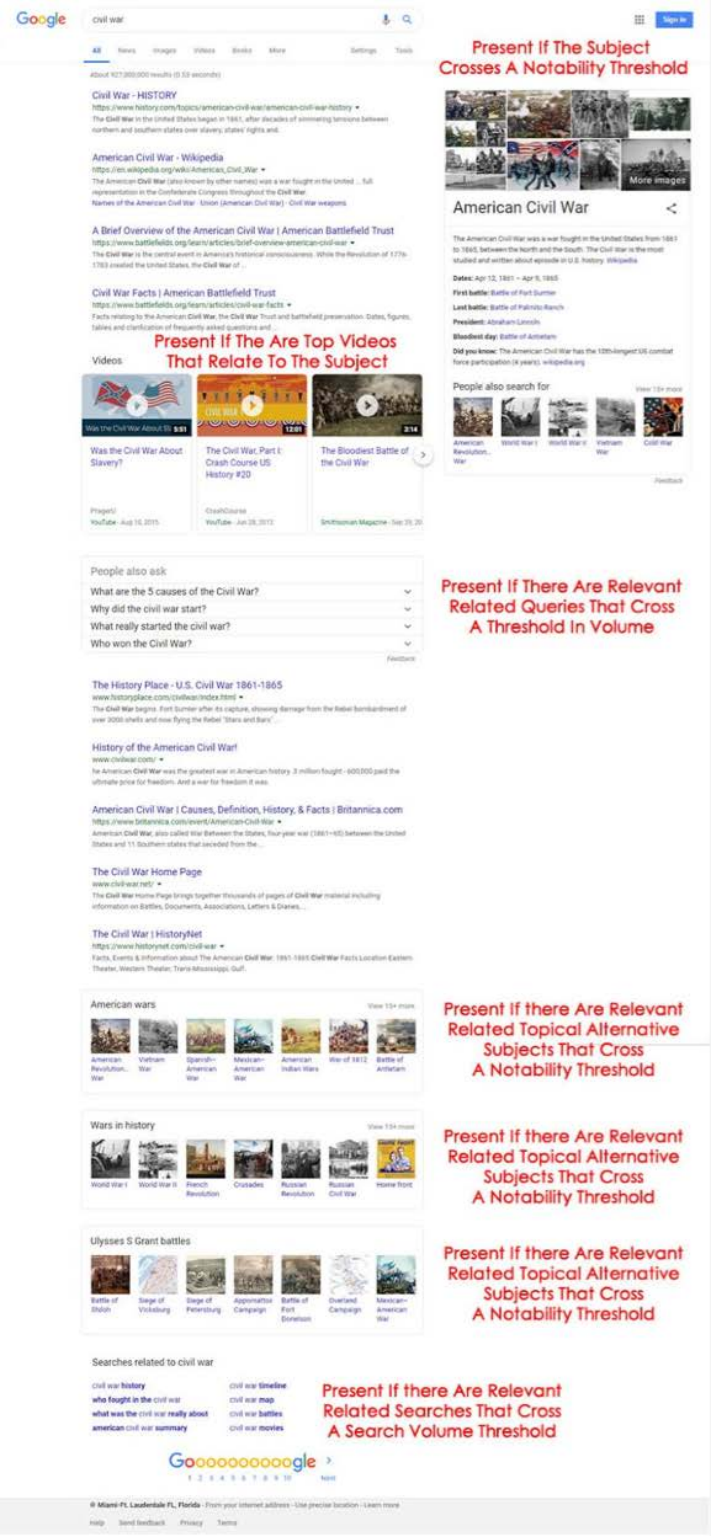
STEP 4

LAYOUT

We've all seen it. In fact, you can see it in the civil war example above. For different queries the search results page layout changes.

The engines will determine what possible formats apply to a query intent, the user running the query and the available resources.

The full page of the SERP for [civil war] looks like:



I've put an educated guess on the core factor used to determine when each element is present.

The truth is, it's a moving target and relies on a knowledge of **entities**, how they connect, and how they are weighted. That's a highly complex subject so we won't dive into it here.

What's important to understand in the context of this piece is that the different elements of any given search results page need to be determined more-or-less on the fly.

This is to say, when a query is run and the first three steps completed the engine will reference a database of the various possible elements to insert onto the page, the possible placements and then determine which will apply to the specific query.

AN ASIDE: I noted above that the search results pages were generated more-or-less on the fly.

While this would be true of infrequent queries, for common queries it is far more likely that the engines keep a database of which elements they have already calculated to fit the likely user intent so as to not have to process that each time.

I would imagine there is a time limit on it after which it refreshes and I suspect that it refreshes the full entry at time of low use.

But moving on, the engine now knows the classification of a query, the context the information is being requested in, the signal weights that apply to such a query, and the layout most likely to meet the various possible intents for a query.

Finally, it's time for ranking.

STEP 5

RANKING

Interestingly, this is probably the easiest step of the process, though not as singular as one might think.

When we think of organic rankings we think of the 10 blue links. So let's start there and look at the process thus far:

- The user enters a query.

- The engine considers the type of query and classifies it to understand what key criteria apply at a high level based on similar or identical previous query interactions.

• The engine considers the user's position in space and time to consider their likely intents.

• The engine takes the query classifications and user-specific signals and uses this to determine which signals should hold what weights.

• The engine uses the above data to also determine which layouts, formats, and additional data may satisfy or supplement the user's intent.

With all this in hand and with an algorithm already written, the engine needs simply crunch the numbers.

They will pull in the various sites that can be considered for ranking, apply the weights to their algorithms, and crunch the number to determine the order that the sites should appear in the search results.

Of course, they must do this for each element on the page in various ways.

Videos, stories, entities, and information all change, so the engines need to order not just the blue links but everything else on the page as well.

IN SHORT

The ranking of the site is easy. It's putting everything together to do it that's the real work.

You may ask how understanding this can help you with your SEO efforts. It's like understanding the core functions of how your computer works.

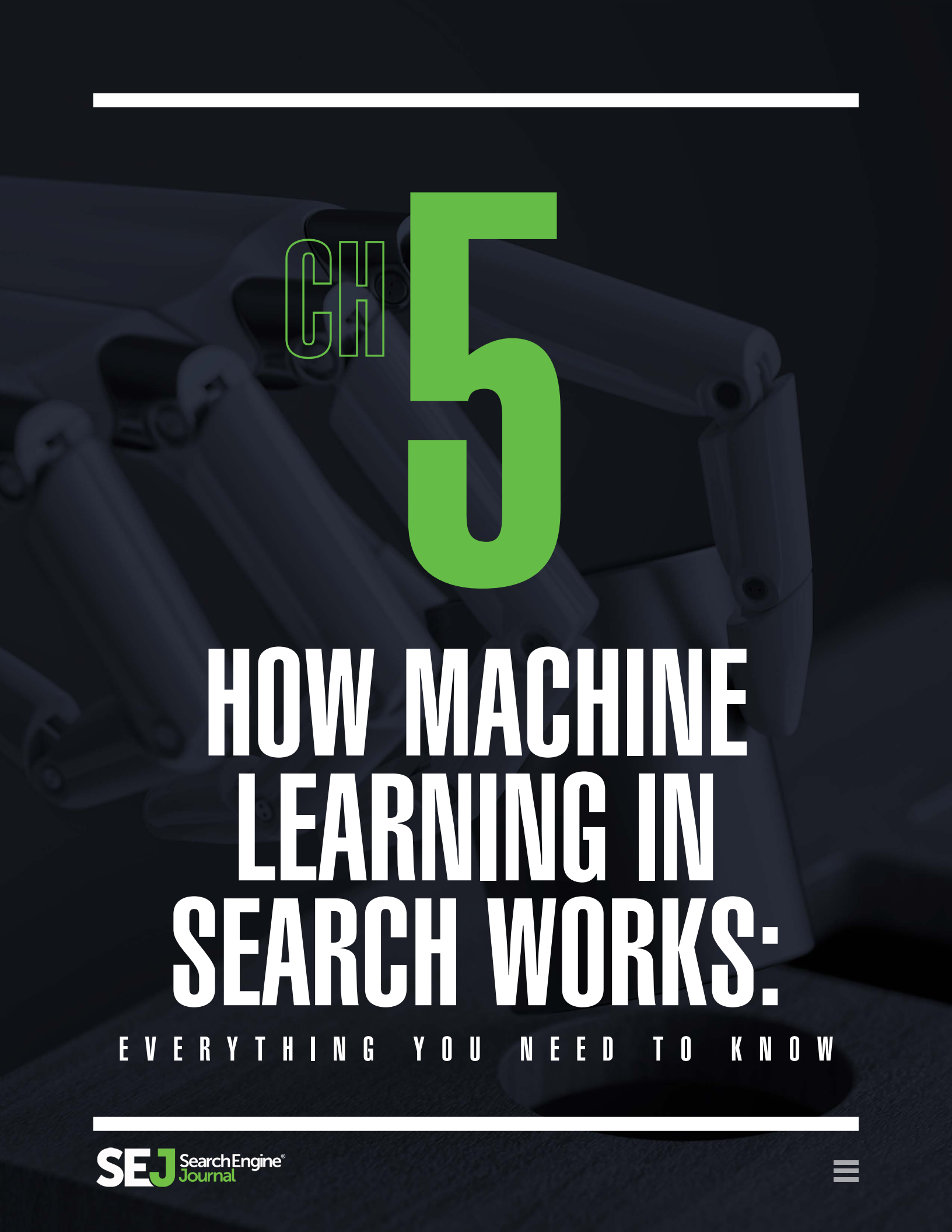
I can't build a processor, but I know what they do, and I know what characteristics make for a faster one and how cooling impacts them.

Knowing this results in me having a faster machine that I need to update and upgrade far less often.

The same is true for SEO.

If you understand the core of how the engine function you will understand your place in that ecosystem.

And that will result in strategies designed with the engine in mind and serving the real user – their user.

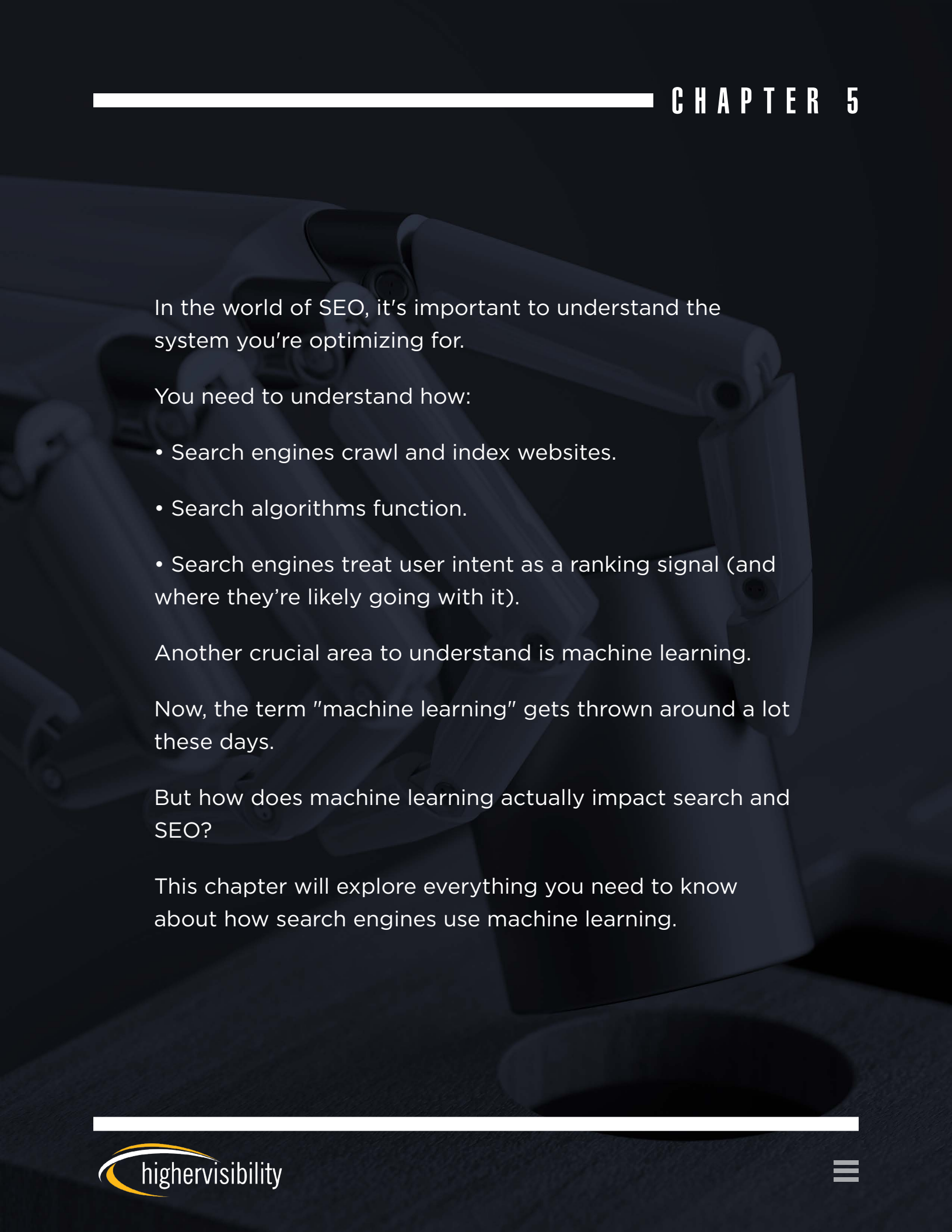


CH 5

HOW MACHINE LEARNING IN SEARCH WORKS:

EVERYTHING YOU NEED TO KNOW

SEJ Search Engine Journal



In the world of SEO, it's important to understand the system you're optimizing for.

You need to understand how:

• Search engines crawl and index websites.

• Search algorithms function.

• Search engines treat user intent as a ranking signal (and where they're likely going with it).

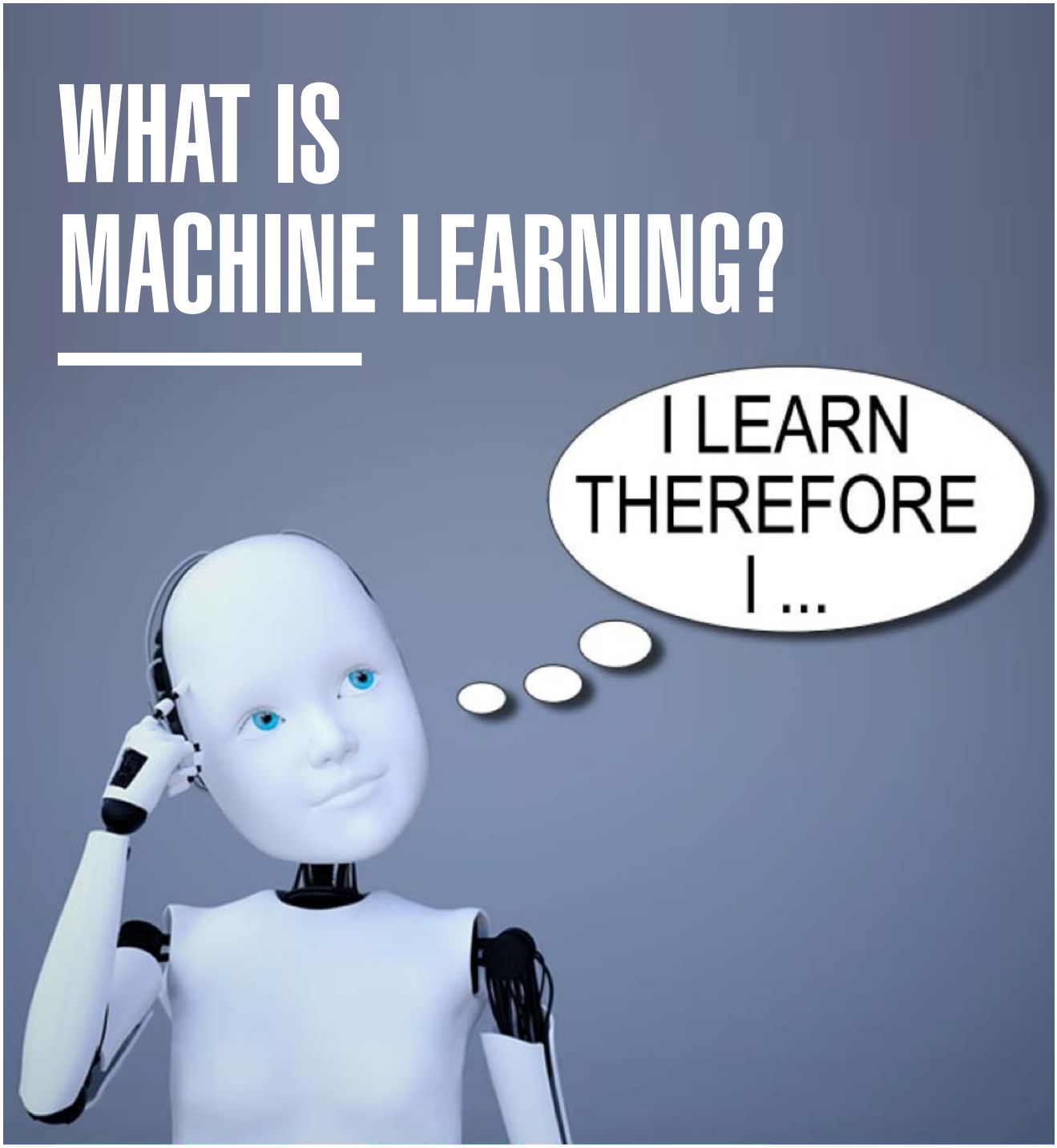
Another crucial area to understand is machine learning.

Now, the term "machine learning" gets thrown around a lot these days.

But how does machine learning actually impact search and SEO?

This chapter will explore everything you need to know about how search engines use machine learning.

WHAT IS MACHINE LEARNING?



It would be difficult to understand how search engines use machine learning without knowing what machine learning actually is.

Let's start with the definition (provided by Stanford University in their course description for Coursera) before we move on to a practical explanation:

"MACHINE LEARNING IS THE SCIENCE OF GETTING COMPUTERS TO ACT WITHOUT BEING EXPLICITLY PROGRAMMED."

A QUICK ASIDE BEFORE WE CONTINUE...

Machine learning isn't the same as Artificial Intelligence (AI), but the line is starting to get a bit blurry with the applications.

As noted above, machine learning is the science of getting computers to come to conclusions based on information but without being specifically programmed in how to accomplish said task.

AI, on the other hand, is the science behind creating systems that either have, or appear to possess, human-like intelligence and process information in a similar manner.

Think of the difference this way:

Machine learning is a system designed to solve a problem. It works mathematically to produce the solution. The solution could be programmed specifically, or worked out by humans manually, but without this need, the solutions come much faster.

A good example would be setting a machine off to pour through oodles of data outlining tumor size and location without programming in what it's looking for. The machine would be given a list of known benign and malignant conclusions.

With this, we would then ask the system to produce a predictive model for future encounters with tumors to generate odds in advance as to which it is based on the data analyzed.

This is purely mathematical.

A few hundred mathematicians could do this – but it would take them many years (assuming a very large database) and hopefully, none of them would make any errors. Or, this same task could be accomplished with machine learning – in far less time.

When I'm thinking of Artificial Intelligence, on the other hand, that's when I start to think of a system that touches on the creative and thus becomes less predictable.

An artificial intelligence set on the same task may simply reference documents on the subject and pull conclusions from previous studies.

Or it may add new data into the mix.

Or may start working on a new system of electrical engine, foregoing the initial task.

It probably won't get distracted on Facebook, but you get where I'm going.

The key word is intelligence.

While artificial, to meet the criteria it would have to be real thus producing variables and unknowns akin to what we encounter when we interact with others around us.

CHAPTER 5

BACK TO MACHINE LEARNING & SEARCH ENGINES

Right now what the search engines (and most scientists) are pushing to evolve is machine learning.

Google has a free **course on it**, has made its machine learning framework **TensorFlow** open source, and is making **big investments in hardware to run it.**

Basically, this is the future so it's best to understand it.

While we can't possibly list (or even know) every application of machine learning going on over at the Googleplex, let's look at a couple of known examples:

RANKBRAIN

What article on machine learning at Google would be complete without mentioning their first and still highly-relevant implementation of a machine learning algorithm into search?

That's right ... we're talking **RankBrain.**

Essentially the system was armed only with an understanding of entities (a thing or concept that is singular, unique, well-defined and distinguishable) and tasked with producing an understanding of how those entities connect in a query to assist in better understanding the query and a set of known good answers.

These are brutally simplified explanations of both entities and RankBrain but it serves our purposes here.

So, Google gave the system some data (queries) and likely a set of known entities.

I'm going to guess on the next process but logically the system would then be tasked with training itself based on the seed set of entities on how to recognize unknown entities it encounters.

The system would be pretty useless if it wasn't able to understand a new movie name, date, etc.

Once the system had that process down and was producing satisfactory results they would have then tasked it with teaching itself how to understand the relationships between entities and what data is being implied or directly requested and seek out appropriate results in the index.

This system solves many problems that plagued Google.

The requirement to include keywords like "How do I replace my S7 screen" on a page about replacing one should not be necessary.

You also shouldn't have to include "fix" if you've included "replace" as, in this context, they generally imply the same thing.

CHAPTER 5

RankBrain uses machine learning to:

• Continuously learn about the connectedness of entities and their relationships.

• Understand when words are synonyms and when they are not (replace and fix may be synonyms in this case but they wouldn't be if I was querying "how to fix my car").

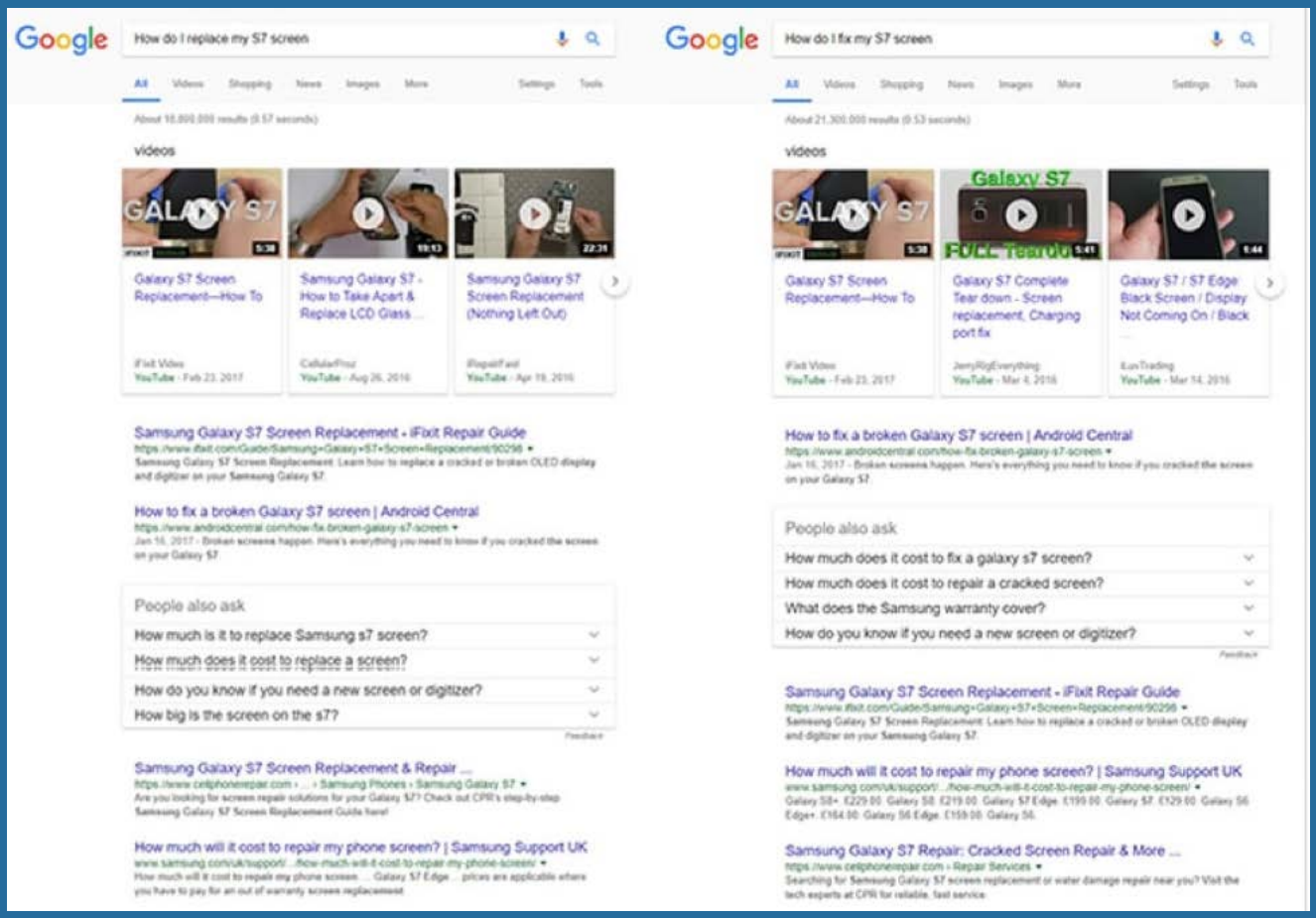
• Instruct other portions of the algorithm to produce the correct SERP.

In its first iteration, RankBrain was tested on queries Google had not encountered before. This makes perfect sense and is a great test.

If RankBrain can improve results for queries that likely weren't optimized for and will involve a mix of old and new entities and services a grouping of users who were likely getting lackluster results to begin with then it should be deployed globally.

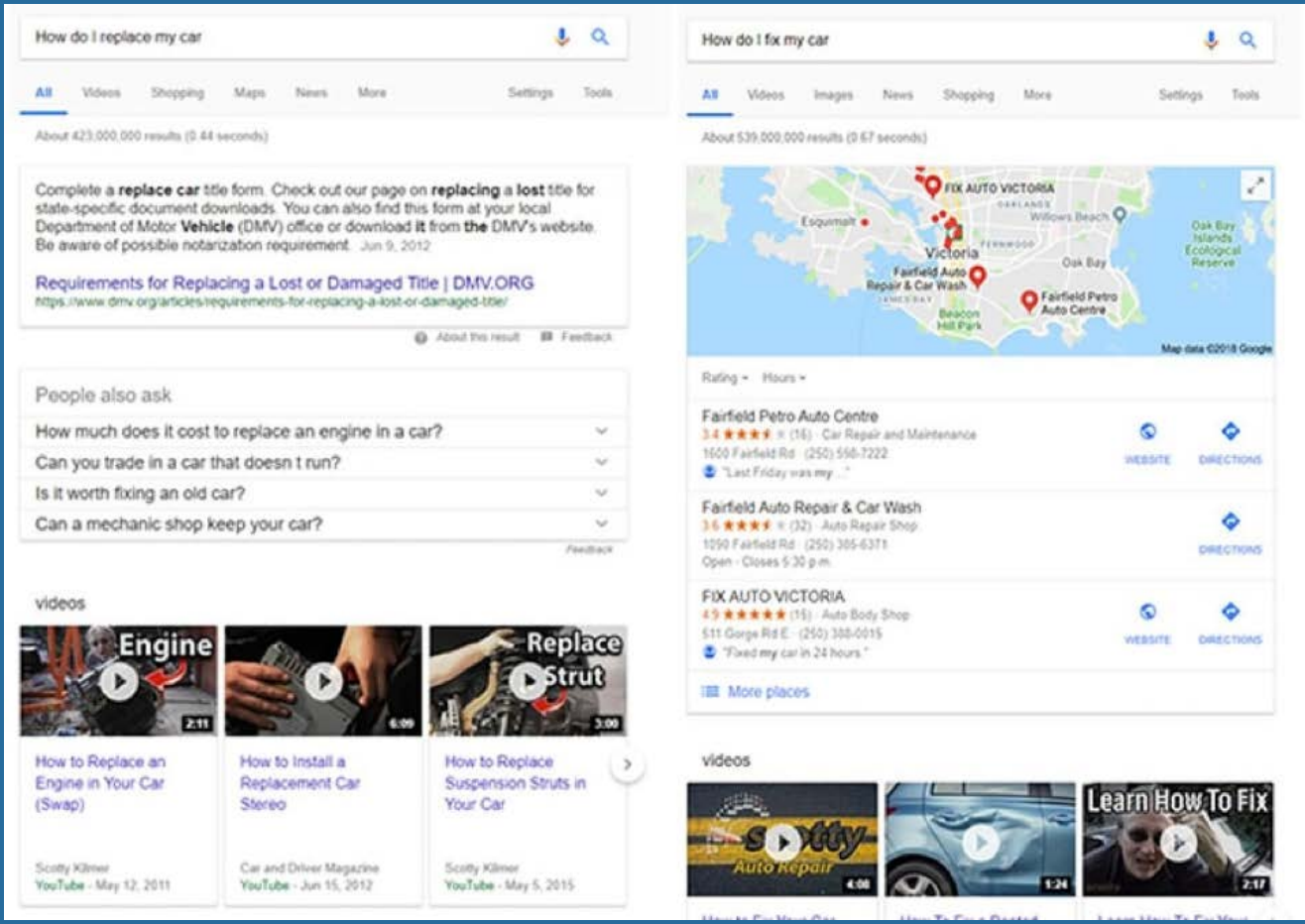
And **it was** in 2016.

Let's take a look at the two results I referenced above (and worth noting, I was writing the piece and the example and then thought to get the screen capture – this is simply how it works and try it yourself … it works in almost all cases where different wording implies the same thing):



Some very subtle differences in rankings with the #1 and 2 sites switching places but at its core it's the same result.

Now let's look at my automotive example:



Machine learning helps Google to not just understand where there are similarities in queries, but we can also see it determining that if I need my car fixed I may need a mechanic (good call Google), whereas for replacing it I may be referring to parts or in need of governmental documentation to replace the entire thing.

We can also see here where machine learning hasn't quite figured it all out.

When I ask it how to replace my car, I likely mean the whole thing or I'd have listed the part I wanted. But it'll learn … it's still in its infancy. Also, I'm Canadian, so the DMV doesn't really apply.

So here we've seen an example of machine learning at play in determining query meaning, SERP layout, and possible necessary courses of action to fulfill my intent.

Not all of that is RankBrain, but it's all machine learning.

SPAM

If you use Gmail, or pretty much any other email system, you also are seeing machine learning at work.

According to Google, they are now blocking 99.9% of all spam and phishing emails with a false-positive rate of only 0.05%.

They're doing this using the same core technique – give the machine learning system some data and let it go.

If one was to manually program in all the permutations that would yield a 99.9% success rate in spam filtering and adjust on the fly for new techniques it would be an onerous task if at all possible.

When they did things this way they sat at a 97% success rate with 1% of false positive (meaning 1% of your real messages were sent to the spam folder – unacceptable if it was important).

Enter machine learning – set it up with all the spam messages you can positively confirm, let it build a model around what similarities they have, enter in some new messages and give it a reward for successfully selecting spam messages on its own and over time (and not a lot of it) it will learn far more signals and react far faster than a human ever could.

Set it to watch for user interactions with new email structures and when it learns that there is a new spam technique being used, add it to the mix and filter not just those emails but emails using similar techniques to the spam folder.

SO HOW DOES MACHINE LEARNING WORK?

This chapter promised to be an explanation of machine learning, not just a list of examples. The examples, however, were necessary to illustrate a fairly easy-to-explain model.

Let's not confuse this with easy to build, just simple in what we need to know.

A common machine learning model follows the following sequence:

• Give the system a set of known data. That is, a set of data with a large array of possible variables connected to a known positive or negative result. This is used to train the system and give it a starting point. Basically, it now understands how to recognize and weigh factors based on the past data to produce a positive result.

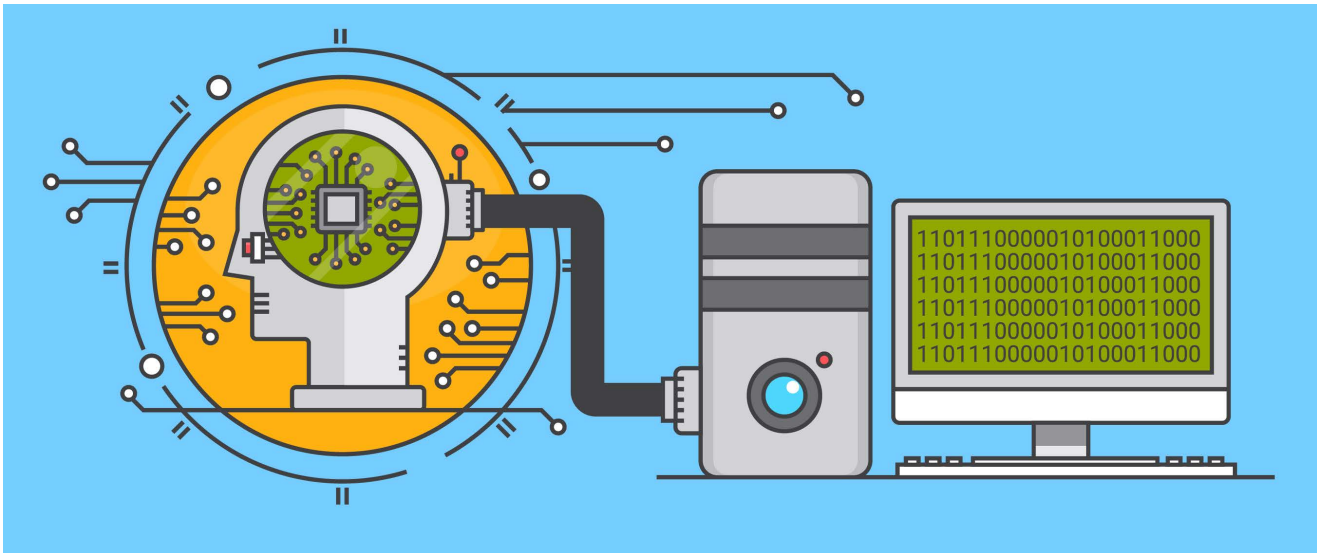
• Set up a reward for success. Once the system is conditioned with the starting data it is then fed new data but without the known positive or negative result. The system does not know the relationships of a new entity or whether an email is spam or not. When it selects correctly it is given a reward though clearly not a chocolate bar. An example would be to give the system a reward value with the goal of hitting the highest number possible. Each time it selects the right answer this score is added to.

•Set it loose. Once the success metrics are high enough to surpass existing systems or meet another threshold the machine learning system can be integrated with the algorithm as a whole.

This model is referred to as supervised learning and if my guess is right, it's the model used in the majority of the Google algorithm implementations.

Another model of machine learning is the Unsupervised Model. To draw from the example used in a great course **over on Coursera** on machine learning, this is the model used to group similar stories in Google News and one can infer that it's used in other places like the identification and grouping of images containing the same or similar people in Google Images.

In this model, the system is not told what it's looking for but rather simply instructed to group entities (an image, article, etc.) into groups by similar traits (the entities they contain, keywords, relationships, authors, etc.)



WHY DOES THIS MATTER?

Understanding what machine learning is will be crucial if you seek to understand why and how SERPs are laid out and why pages rank where they do.

It's one thing to understand an algorithmic factor – which is an important thing to be sure – but understanding the system in which those factors are weighted is of equal, if not greater, importance.

For example, if I was working for a company that sold cars I would pay specific attention to the lack of usable, relevant information in the SERP results to the query illustrated above.

The result is clearly not a success. Discover what content would be a success and generate it.

Pay attention to the types of content that Google feels may fulfill a user's intent (post, image, news, video, shopping, featured snippet, etc.) and work to provide it.

I like to think of machine learning and its evolution equivalent to having a Google engineer sitting behind every searcher, adjusting what they see and how they see it before it is sent to their device.

But better – that engineer is connected like the Borg to every other engineer learning from global rules.

But we'll get more into that in our next piece on user intent.

CH 6

HOW USER BEHAVIOR IN SEARCH WORKS:

EVERYTHING YOU NEED TO KNOW

In this age of search, it's not enough to know what ranking signals may-or-may-not be at play.

You also need to understand the environment in which those signals operate.

This chapter will look at how user behavior is – and isn't – used by search engines. Unlike the previous three installments, which focused on principles, this post will rely on more specific examples and applications.

Most of this chapter will discuss Google because they are the leader in this space, but the principles also can be applied to Bing.

USER BEHAVIOR SIGNALS THE ENGINES DON'T USE

Before we dive into how user behavior impacts search, let's make sure we're all clear on what user behavior does not impact.

GOOGLE ANALYTICS

Having Google Analytics on your site does not impact search one way or the other.

When asked about whether there was a penalty for removing Google Analytics on June 28, 2018, Google's John Mueller **responded:**

Now, I'm not saying that you should take everything Google says at face value, but this is a repeated response and it makes sense.

Not all website owners/webmasters/SEO professionals use Google Analytics. If Google used their analytics platform to determine success and failure metrics of your site specifically they would essentially be comparing apples-to-oranges.

Your individual user metrics (as reported or detected by Google Analytics) are not used specifically for or against you.

SOCIAL MEDIA LIKES, SHARES & COMMENTS

The notion that the number of likes, shares, followers, etc. on Facebook (or Twitter, or Instagram, or any other social media platform) has been refuted repeatedly by Google. Again, this makes perfect sense.

Google has clearly stated that social signals are:

• Too easy to manipulate.

• Unreliable because it's impossible to consistently get all the available content/data from these platforms.

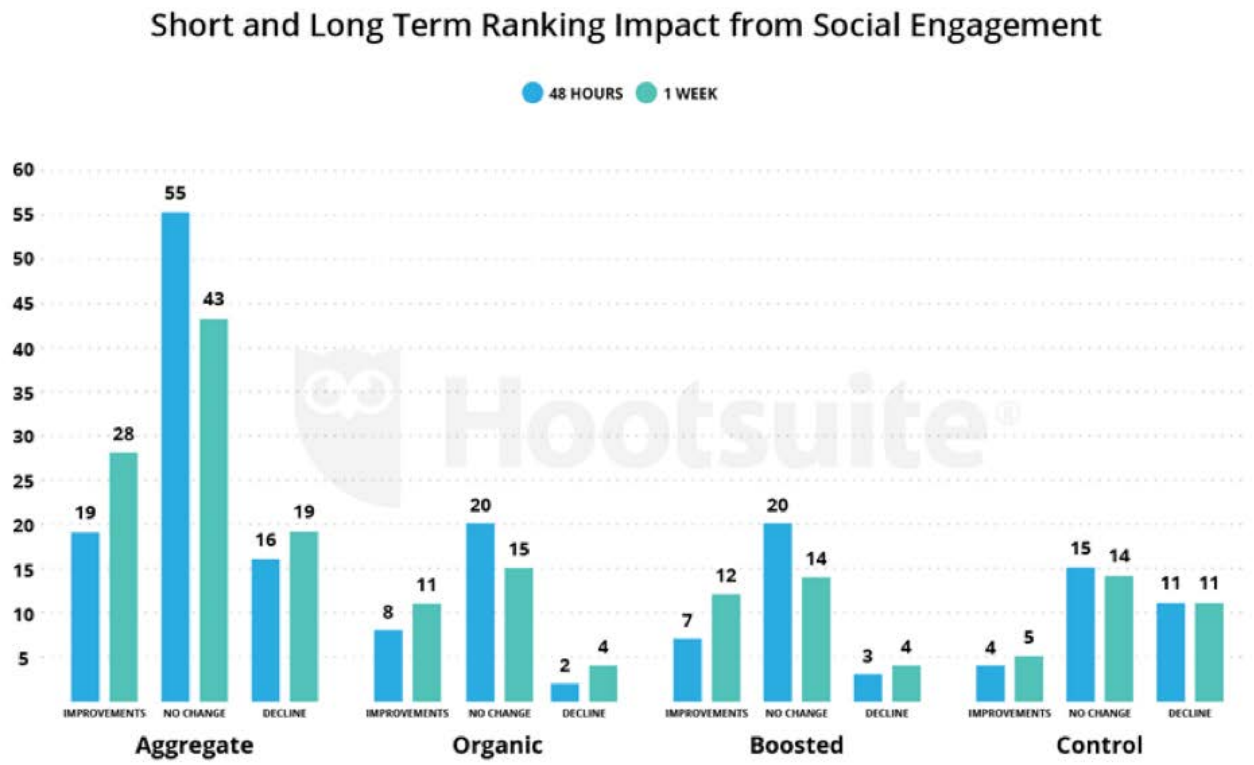
To be balanced, I do need to reference a Hootsuite **study** from a couple of months ago that found a correlation. This study was limited to Twitter and 90 pages of content divided into three groups:

• No promotion: The control group. No organic tweets or paid promotion for 30 articles.

• Organic promotion: Thirty articles were promoted via organic tweets.

• Paid promotion: The remaining 30 articles were organic tweets that were then boosted.

The result was:



Basically, both sets of pages tweeted showed stronger ranking improvements in timeframes that were too short to be attributable to links.

I bring this up not to confuse the issue but to provide full data.

Google has said they don't use it as a metric and this is likely true. So how can it be that the evidence above illustrates the opposite?

First, the data is not conclusive at its scale, which is mentioned in Hootsuite's article. It's more likely that what we're seeing here isn't Twitter impacting a ranking, but rather the reinforcing of entity relevancy.

Hashtags and keywords tied to an **entity** and then being associated with the page referenced in the tweet.

To this end, it's not that it was Twitter or any other site, it's simply that we're seeing a non-link-based entity reinforcement resulting in a ranking improvement. Something that matches well with everything we're seeing in patents and hearing from Google.

USER BEHAVIOR SIGNALS THAT SEARCH ENGINES CAN USE

Now that we've covered what user signals the engines don't use (or at least some of the core aspects I hear discussed on a regular basis), let's look at the type of user behavior signals that search engines can use.

CLICKS & POST-CLICK BEHAVIOR

While Google has stated correctly that they don't use Google Analytics to gather signals to use in ranking a page that doesn't mean they don't use post-click signals at all.

Google knows:

- Which sites users click on in the search results.
- How long a user was at the target site before they returned to Google.
- What users did next.

SO ESSENTIALLY, WE HAVE FOUR CORE POSSIBILITIES GOOGLE CAN USE:

1. A searcher clicks to your site in the search results and quickly returns to the engine and clicks on the next link. This would be a clear indication that the searcher believes the query is right and the result wrong and send a negative relevancy signal to the engine for your page as it relates to that query.

2. A searcher clicks to your site in the search results, sticks around for a good deal of time, and then returns to the engine and clicks on the next link for the same query. This would likely send a positive relevancy signal to Google for that query and indicate that the user found content of interest and is simply looking for additional information or options.

3. A searcher clicks to your site in the search results and after either a long or short period of time returns to the engine, refines their query, and searches again. This would indicate to the engine that the user themselves may not have been as specific as they needed to be and not apply a good or bad relevancy to your site.

4. A searcher clicks to your site in the search results and, after either a long or short period of time, returns to the engine and completely changes their query. This would indicate that they found what they were looking for and moved on to a different task. This would pass a positive relevancy with this signal.

While these signals will obviously depend on factors such as the query type and user behavior patterns, this serves as an illustration of the types of signals that can be gleaned from your click data.

We can expect to see a lot more of this as machine learning systems get more and more sophisticated. I wouldn't be surprised if it became a core signal for ranking sites as little else can confirm the success or failure of a SERP placement as well.

REVIEWS

No discussion on user behavior as a ranking signal would be complete without mentioning reviews.

As SEJ's Matt Southern discussed back in November of 2017, reviews are the most prominent signal as it relates to local search. You can read the full article **here.**

This is a direct example of Google using a user behavior signal directly.

There are also patents that discuss the use of location as a signal in the capacity of having run a query for a product or service and then detecting via Android or other systems that you ended up there, thus confirming the fulfillment of your intent.

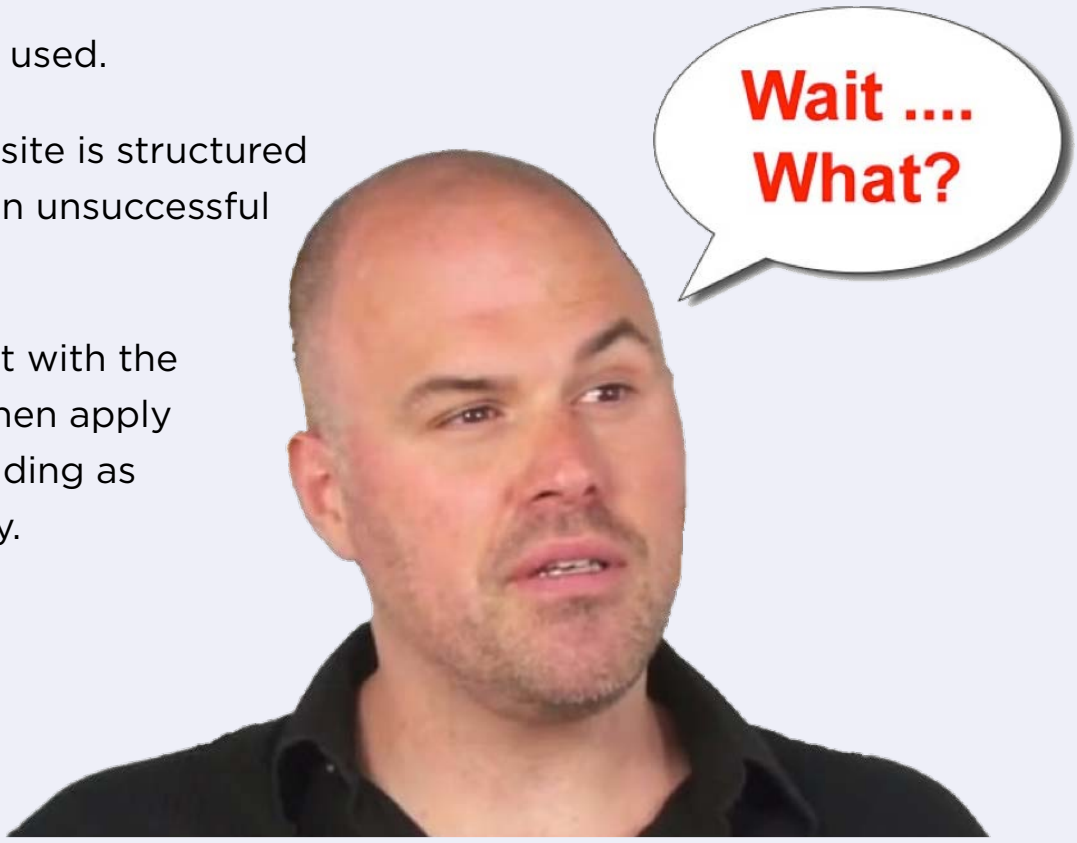
Using your location as a review and taking into account whether you have passed other businesses offering the same service or product and strengthening the value of that signal if you did (i.e., if you'd travel farther to engage with a specific location it must be superior).

GOOGLE ANALYTICS – POSSIBLY

I know I noted above that Google has said they don't use Google Analytics as a ranking signal and that's correct. But that doesn't mean they don't use it to create ranking signals.

Google could use Google Analytics to understand how:

- Websites are used.
- A successful site is structured compared to an unsuccessful one.
- Users interact with the website, and then apply that understanding as signals globally.



To be clear, there is a reason that I put “possibly” in the heading of this section. I am not certain.

From the way things are worded in their denials of the use of Google Analytics, it seems probable.

Even if they aren’t using it today due to complexity, it’s definitely something a little machine learning could make feasible and help Google predict with a higher degree of certainty whether a new webpage or website will satisfy a searcher’s intent.

The purpose of this series has been to lend insight into the approach search engines are taking to search and how they rank pages, not to discuss specific strategies.

In the case of user behavior metrics, this was easiest done will examples. Don't be satisfied with what you've read above, though. They were just examples.

Here's what you must always remember:

THE USER IS THE REASON.
If an engine can understand what a user's intent is, and that the likelihood is of that intent being met by a specific webpage or website, they have grasped the Holy Grail.

We've seen how an engine can (or in some cases simply may) use a visitor's behavior as a search signal, but clearly this is an end goal.

What we need to do now is to understand how every user action could be measured and weighted. If this can be done in a manner that could be applied globally, it will almost certainly be used as a signal in some capacity.

To do this you need to consider:

- What technologies are at play.
- What its limitations are.
- How the signals they provide can be aggregated.
- Whether it either is, or will be, a signal.

The best part?

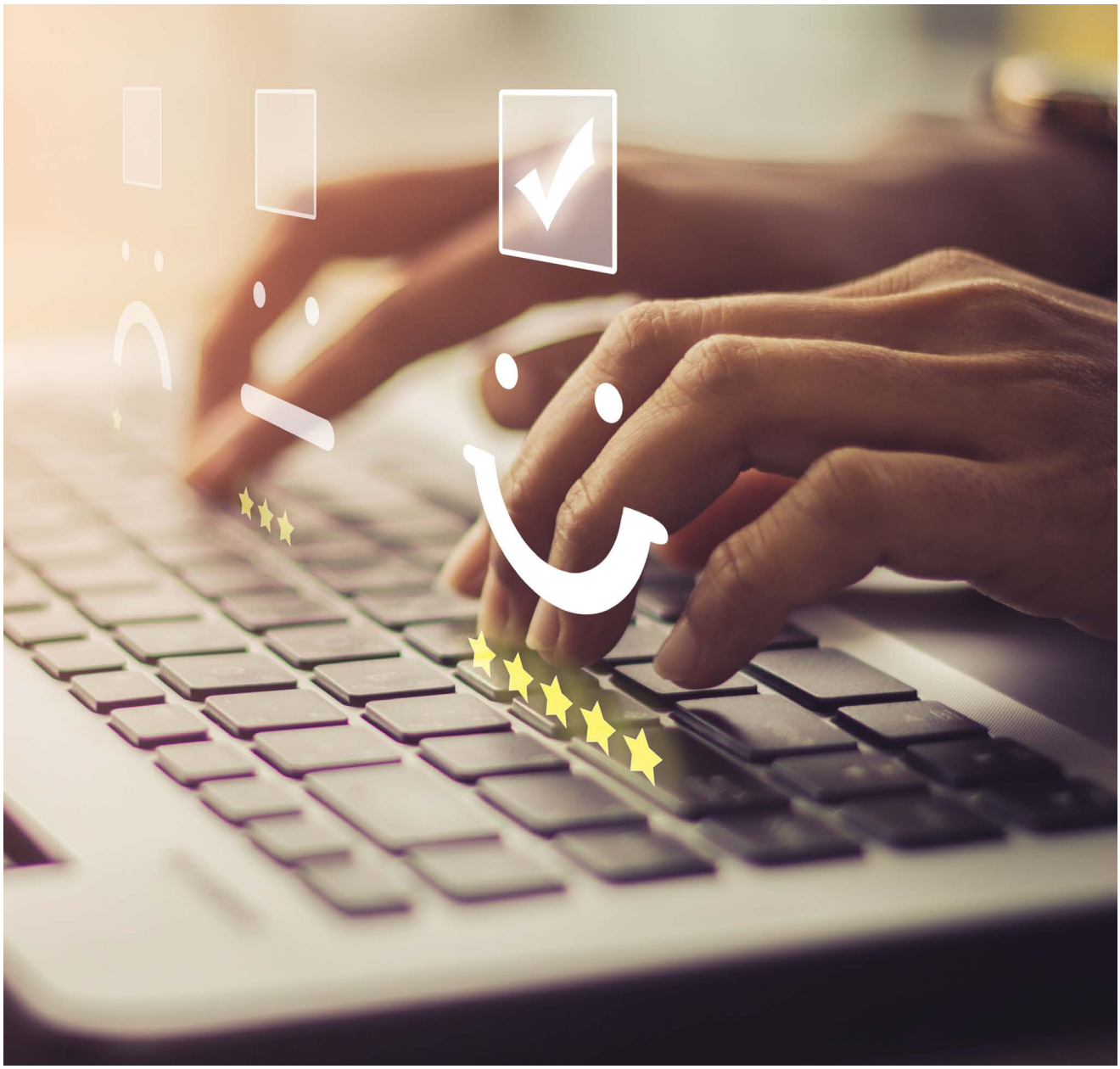
In this quest, there's always a chance you may be wrong.

You may be considering a user behavior that a search engine isn't using.

And in that wrongness, you will be focusing on sending positive user signals and creating the experience that would generate them.

Even if the engine doesn't thank you directly … your bottom line will.

And isn't that the point?



CH 7

HOW SEARCH ENGINES DISPLAY SEARCH RESULTS

Let me begin this with a full disclaimer. I begin each day by ransacking the news to make sure I know what's going on in the search world around me.

Follow me on Twitter and at some point, in the morning you'll find a flurry of Tweets – that's when.

For a slide deck I had put together recently, I decided to publish each change in the SERP (search engine results pages) layouts for the month prior. There were 18 slides in that section. And that was just for February 2019.

I want to stress this point, a point we will come back to later. It's important.

But for now, all we need to keep in mind is that there is a good chance that between the second this piece is published and the time you are reading there may well have been changes.

Actually, there's a very likely chance that between the time I finish writing it, it gets edited, and publishes, there may well have already been changes.

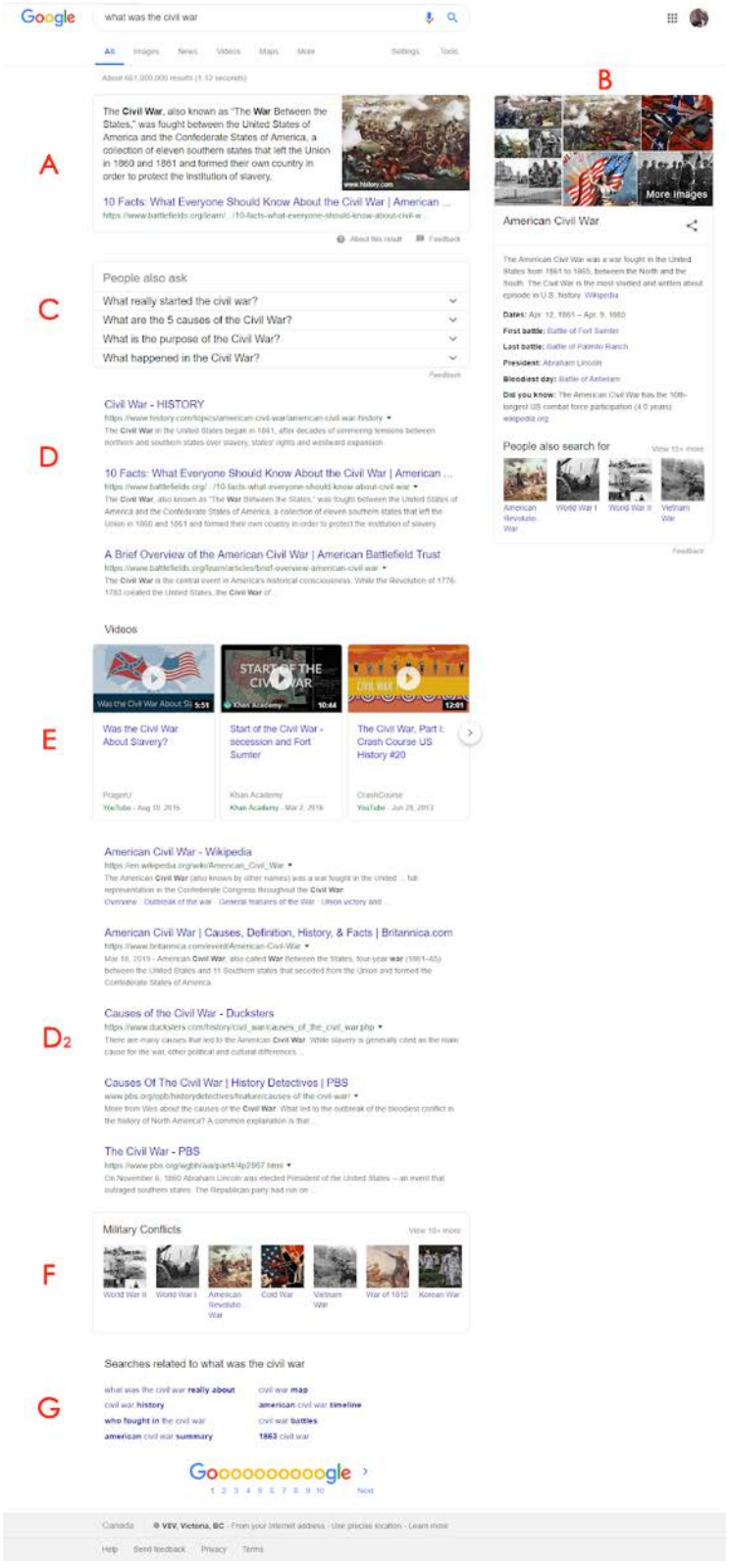
Yes, the pace of change in the SERPs is that fast.

They may not be huge... but they're there and through more than a dozen per month, over a year even that small once create dramatically different experiences.

So, what we will focus on here are the main blocks and some of the elements on them. That is to say, the main areas, where the data is gathered to produce them and what that means for you.

GENERIC SERP LAYOUT

Let's start by looking at a pretty generic SERP layout:



This isn't the only layout as we'll see below but it's likely pretty familiar to you.

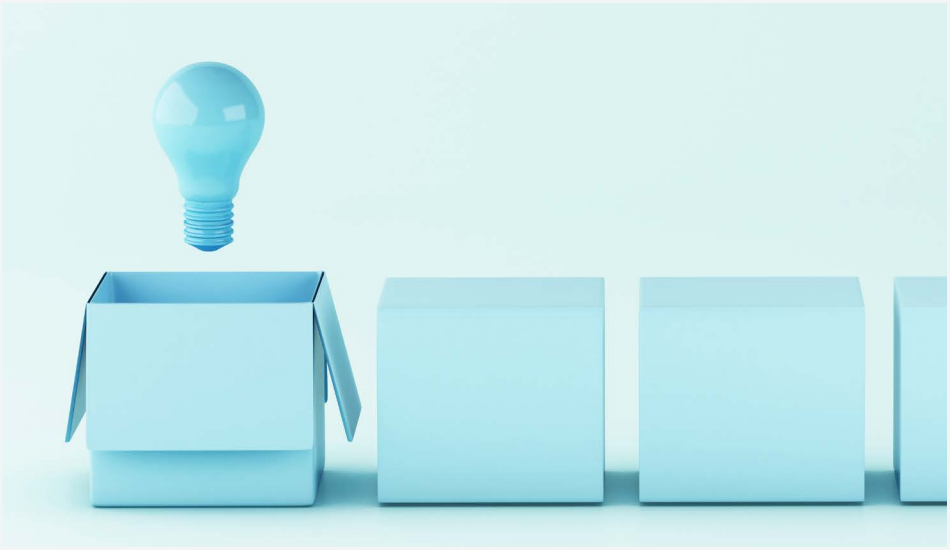
So, what are these sections?

A FEATURED SNIPPET / ANSWER BOX

This is the section above the organic results that attempts to answer a user's complete intent.

As we can see in the example above, if the only intent is a simple answer, this is where it'll likely (though not exclusively) be.

Importantly, structuring your content in a way that produces the answer box often results in the answer for Google voice search as well. But not always... as with the example above. More on that below.



B KNOWLEDGE PANEL / GRAPH

For business or known human **entity** queries, this generally contains a summary of the information Google views as core to their identity. That is, key information a searcher would likely be interested in knowing.

For more general queries, however (like the civil war), we find key facts and images, generally with links to other relevant events or entities.

I noted above that voice search results don't exclusively come from the answer box.

If there is a knowledge panel the voice result will generally come from here. In fact, I've yet to find an exception though it may be a truncated version.

C PEOPLE ALSO ASK

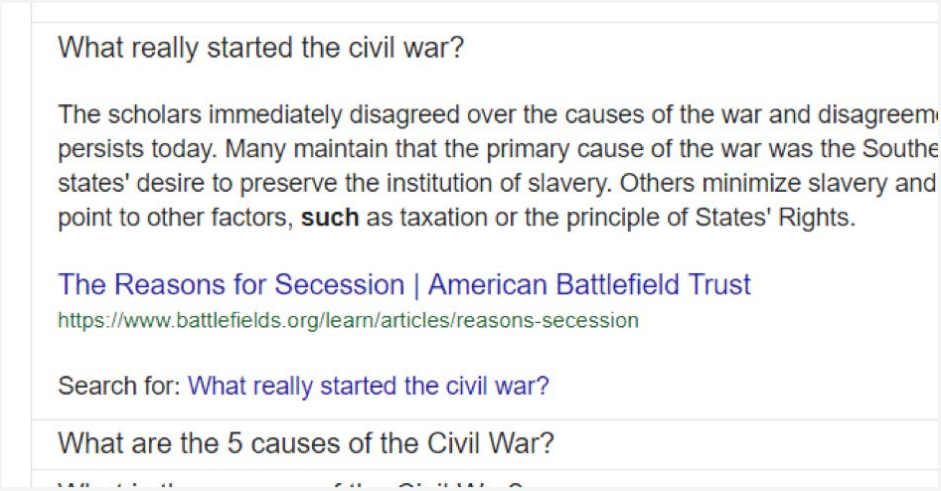
Exactly as the name suggests, this section contains a list of questions that relate to the initial query.

This section is generally triggered when the initial query implies that the user is seeking information on a topic.

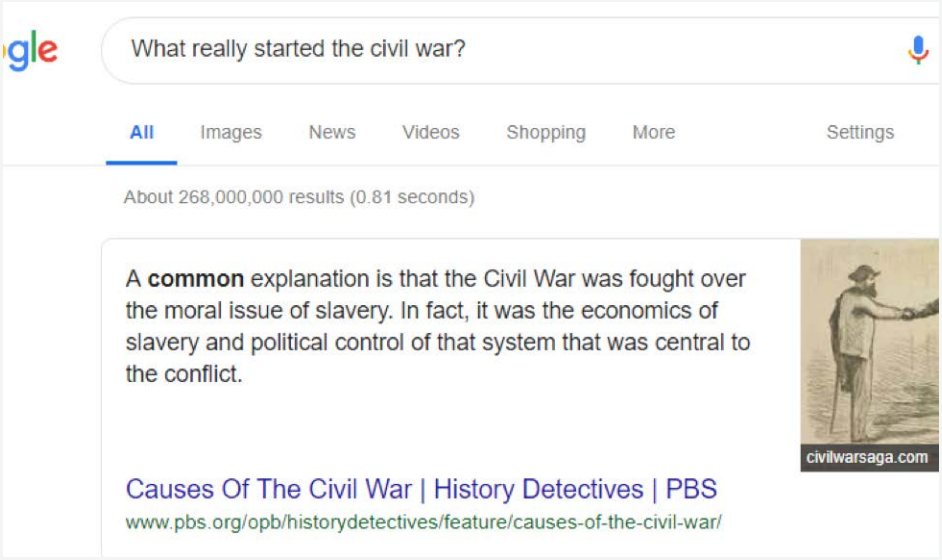
The list of questions relates more to the query itself than search volumes. That is to say, these are not necessarily the top queries around an entity but those questions that relate to the initial question.

When a result is expanded, an answer for the query is given with a link to the site the answer was drawn from as well as a search result for the query with additional details.

Interestingly: The answer given on the initial results page:



Differs from the Answer Box result on the results page if clicked:



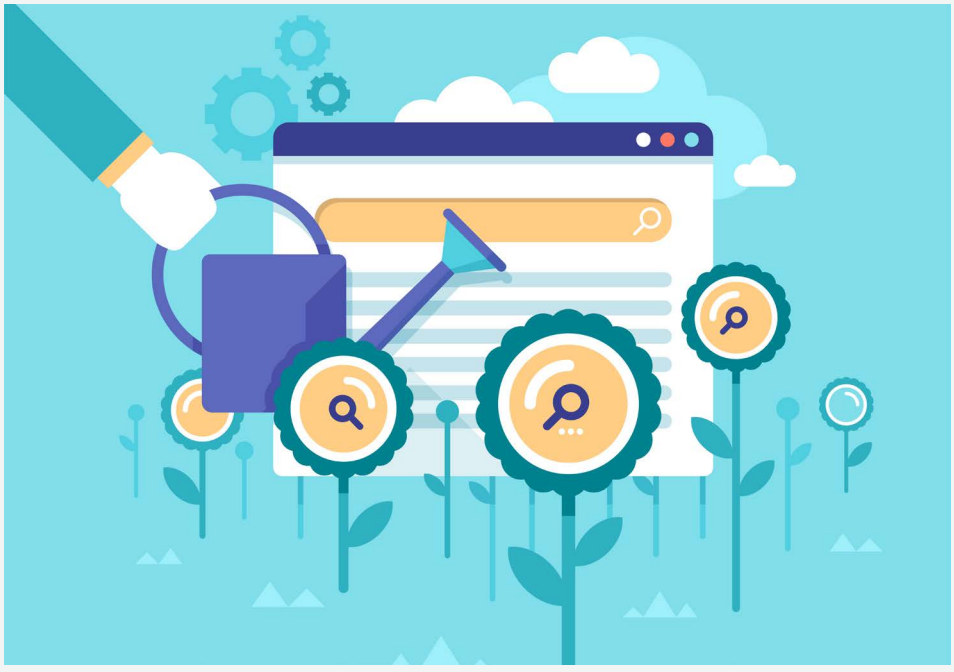
Likely they are assuming that the user's intent differs when the query is being directly searched vs. tacked on to the previous.

D & D2 ORGANIC RESULTS

Technically everything on the page above is an organic result.

As everyone reading this chapter is most certainly aware, these are produced based on a combination of very sophisticated **algorithms** over at the Googleplex(es) and are ordered based on those algorithms – designed to produce the top pages to satisfy a user's likely intent(s).

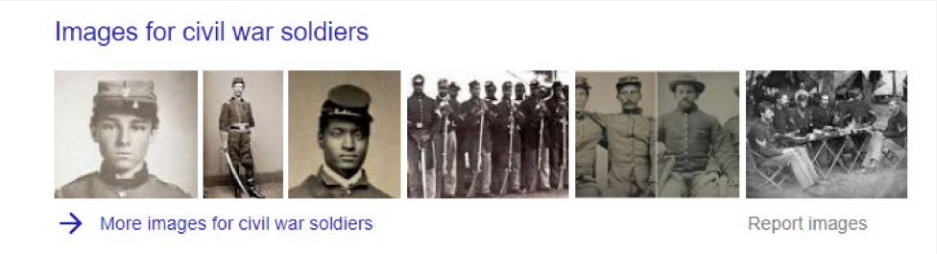
I'm not going to attempt to dive into what signals are used right now as that's not the purpose of this chapter.



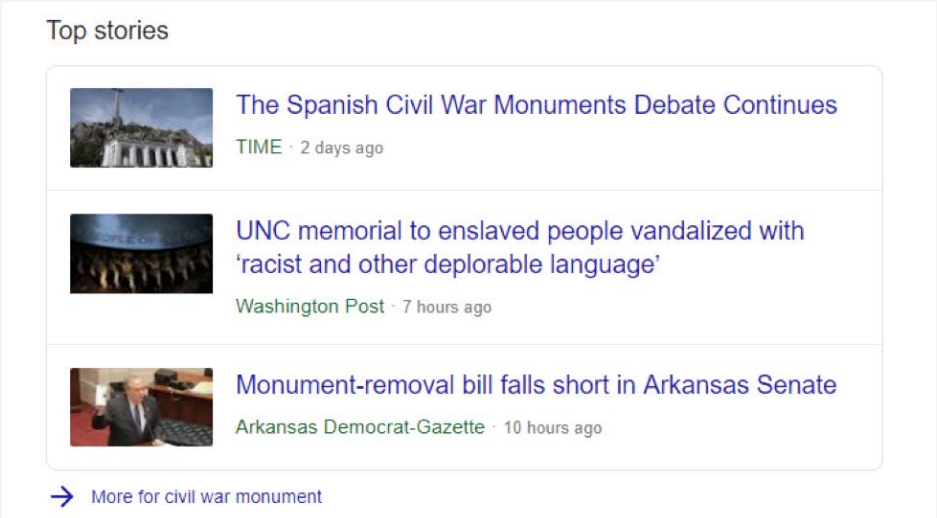
E VIDEO RESULTS (ALTERNATE: NEWS OR IMAGES)

When there are popular videos that attempt to answer a query, they are often displayed in a carousel.

Alternatively, if the query inspires Google to believe that the user intent would be met with the addition of images we'll find:



Or if the query triggers the likely intent that the user may be looking for news:



F RELATED ENTITIES

In section F above we find a row of related entities based on a core characteristic.

In the query used as an example, we were seeking information on a major military conflict.

Google has determined that "military conflict" is the entity association most relevant to the searcher and thus listed others.

There can be more than one such row of results at the bottom of the page though I've yet to see more than three.

G SEARCHES RELATED TO...

At the bottom, we find the related searches.

They differ from the "People Also Ask" in that they don't have to be questions (though they can be). As such, there can be a bit of overlap, but not necessarily.

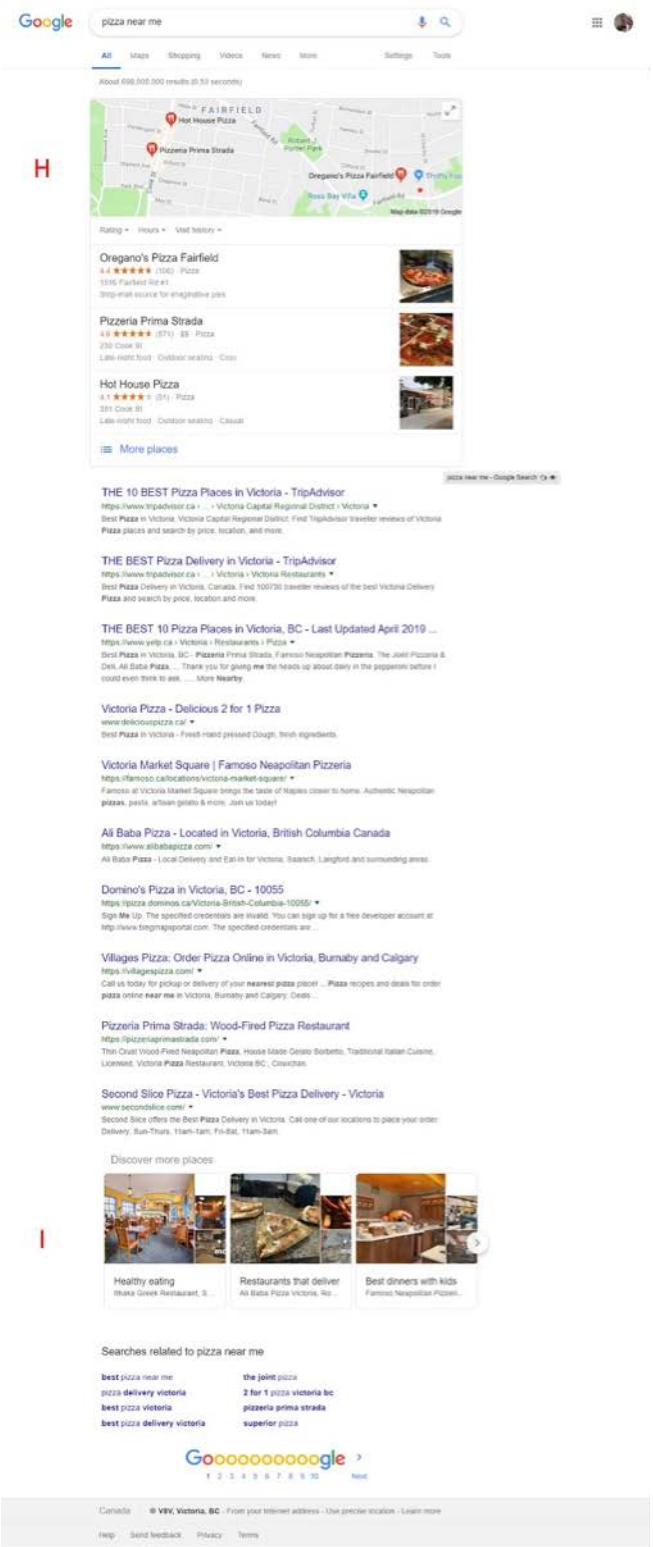
Generally, these are generated by searches that people who searched for the present query have also searched.

LOCAL SERP LAYOUT

Oh, wait... Google hasn't monetized yet and there are some SERP features that are missing.

OK, let's try again.

As it's almost lunch as I write this, let's look up pizza near me. We get:



H SNACK PACK / MAP PACK / LOCAL PACK

For anyone familiar with local in any way or anyone who's ever done any type of query with a local intent, you'll be familiar with the map pack / snack pack / local pack. Wow, that's a lot of names.

Terminology Lesson: For folks newer to SEO, until August of 2015 there were 7 results in the map pack. On August 7, **Google reduced that number to 3.**

As everyone was familiar with 7 being the map pack and this was a far lower number, it became referred to as the snack pack.

If you run a local business and want in the map results, **here's a guide on Local SEO.**

DISCOVER MORE PLACES

This section of the SERPs can be a bit confusing until you really think about it.

- I ran a query for pizza.
- I looked through a variety of results.
- I hit the bottom of the page.
- They're showing me things related to the high-level category but not necessarily related to pizza.

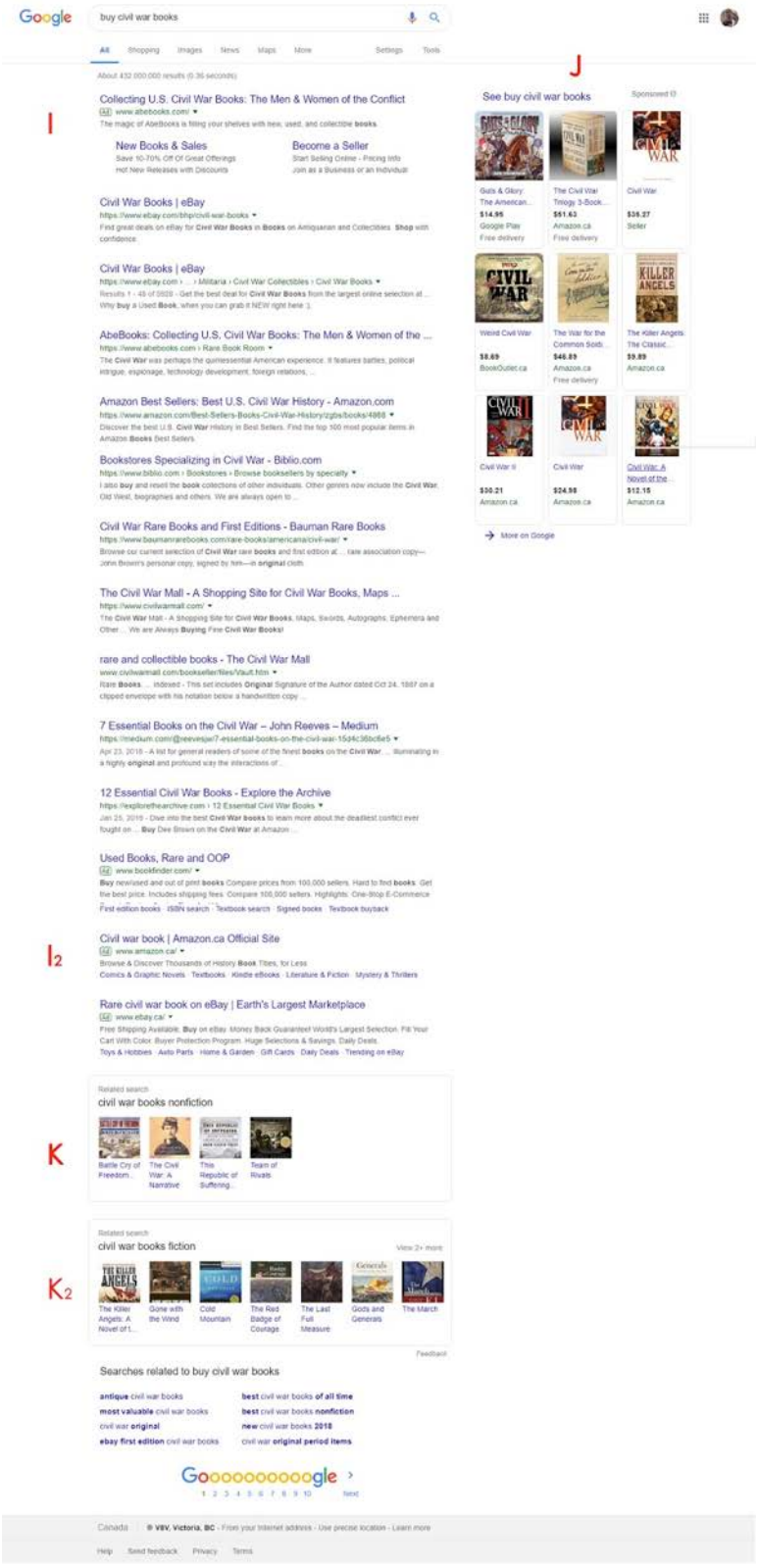
At the bottom of the page, Google has added a section to help me either refine my search, focus it more on sub-categories like delivery, or change gears altogether.

If I hit the bottom of the page, they're assuming I might not have been specific in my desires or even known them and so they're providing new options.

Talk about making page 2 irrelevant.

SERP WITH GOOGLE ADS

Right... all this and we still haven't seen much in the way of ads. So, let's kill two birds with one stone and look at the SERP:



I & I2 ADS

I don't think any of us really need any insight into what this section is for.

It's what pays for all that Google is and let's then do things like **buy Burning Man.**

J SHOPPING RESULTS

Sometimes they're tucked away at the right, sometimes they're placed in a carousel within the results themselves but at its core, the shopping ad units are simply Google Ads power by product-specific data.

If you sell products, have them in a database, invest in Google Ads and don't have a shopping feed set up to power their shopping ads, it's definitely something to look into.

K & K2 RELATED SEARCHES

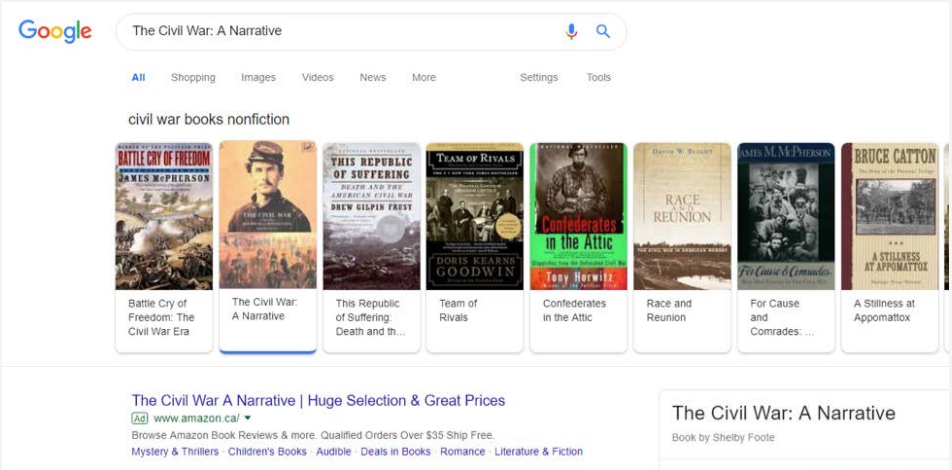
Once again, we see Google dropping a couple of rows of images to distract us from page 2.

These lists are based on entity association on a topical level.

All of the books in the first list relate to the topic of the civil war and the status of being nonfiction. The second list is also related to the topic of the civil war but the status of fiction.

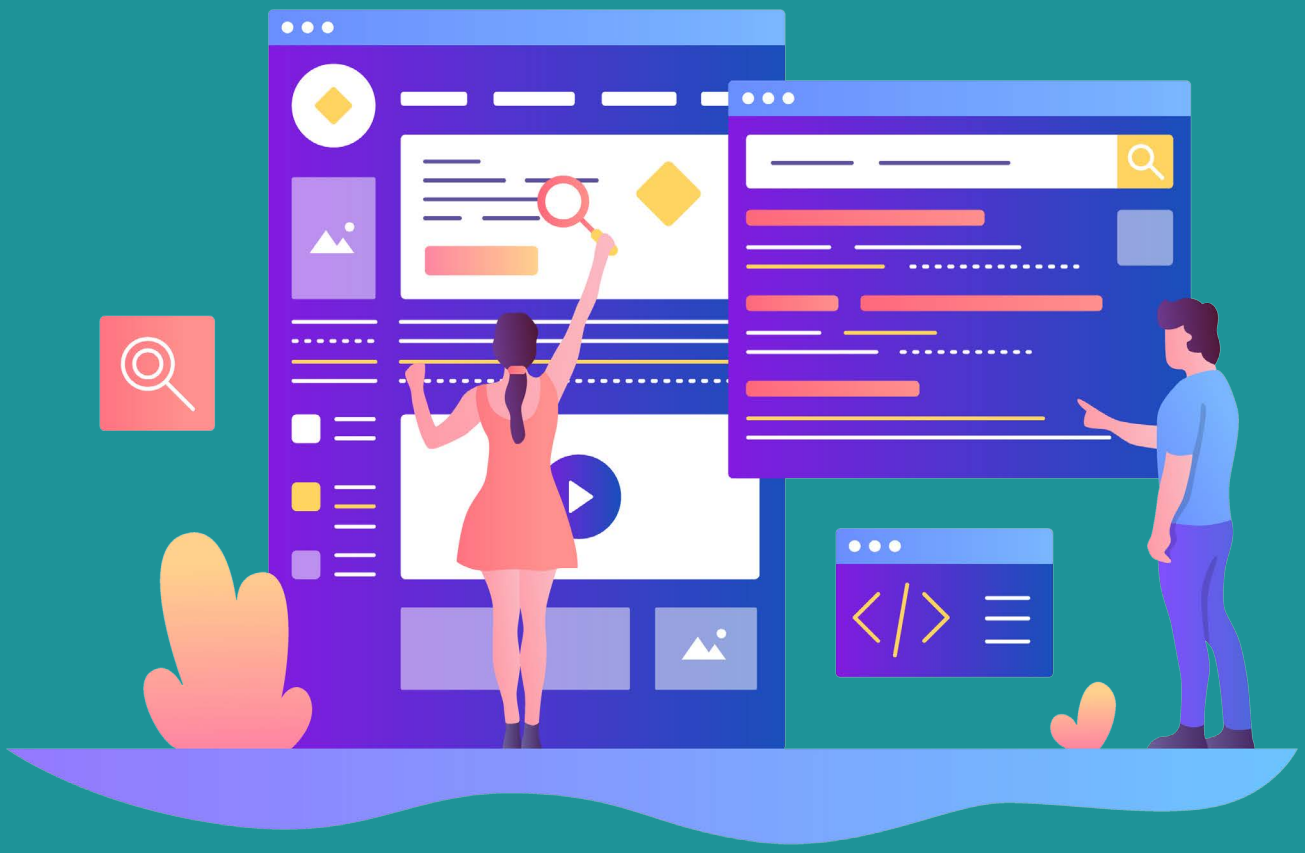
What's interesting is that Google doesn't assume from a click in this zone that you've actually found what you wanted in the first place but rather are inviting you down a different path.

If I click "The Civil War: A Narrative" I am taken to the page:



A carousel at the top displays an expanded version of the list from the previous page. Of course, they take the time to toss in another ad in case I'd like to purchase it.

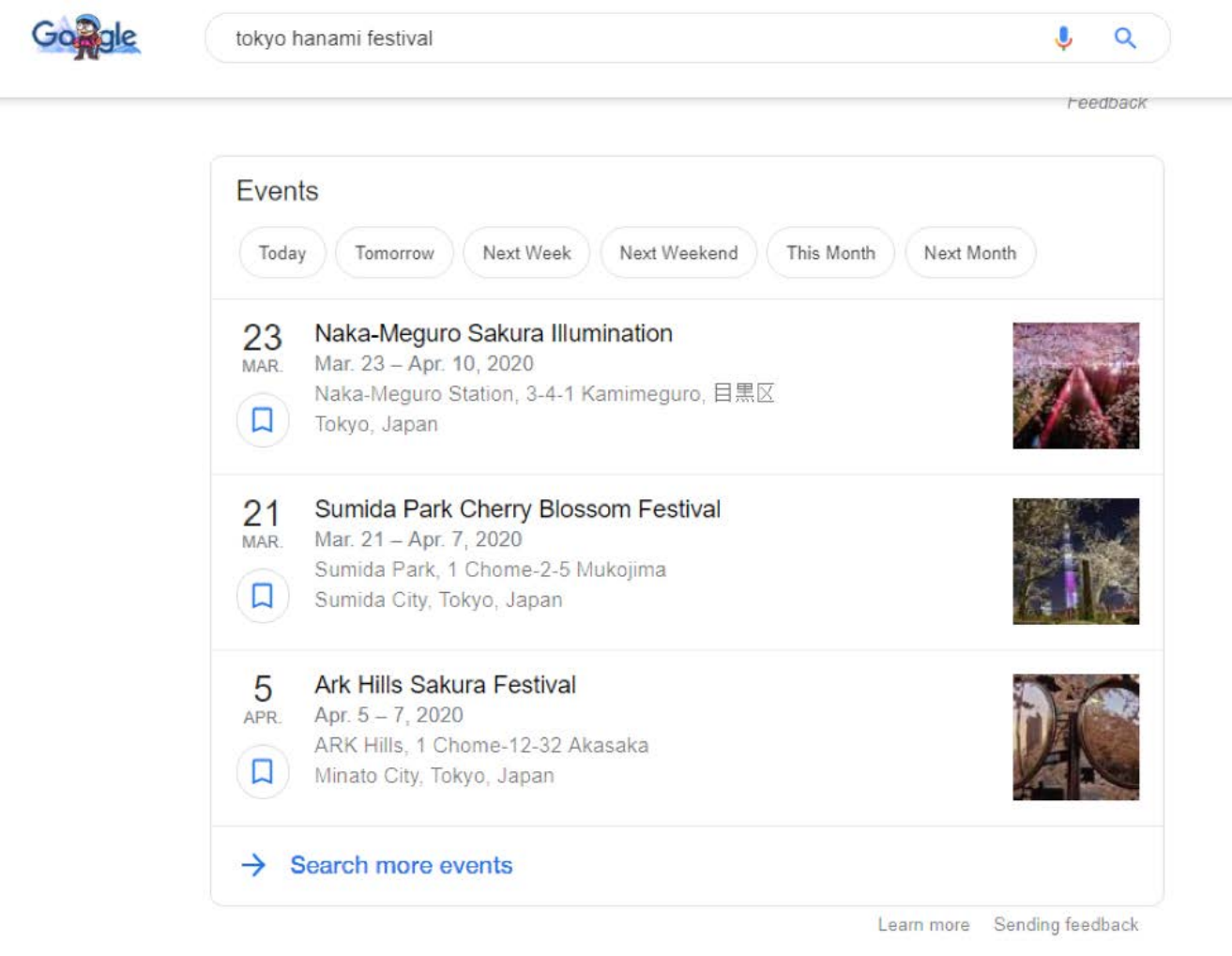
There's a knowledge panel as this is a specifically defined entity and then there are organic results.



ADDITIONAL SERP LAYOUTS & FEATURES

While I will publish this knowing full well that I'm going to miss some due to the sheer volume of different permutations, layouts and sections, here are a few of the more interesting layouts the occupy zones listed above:

EVENTS



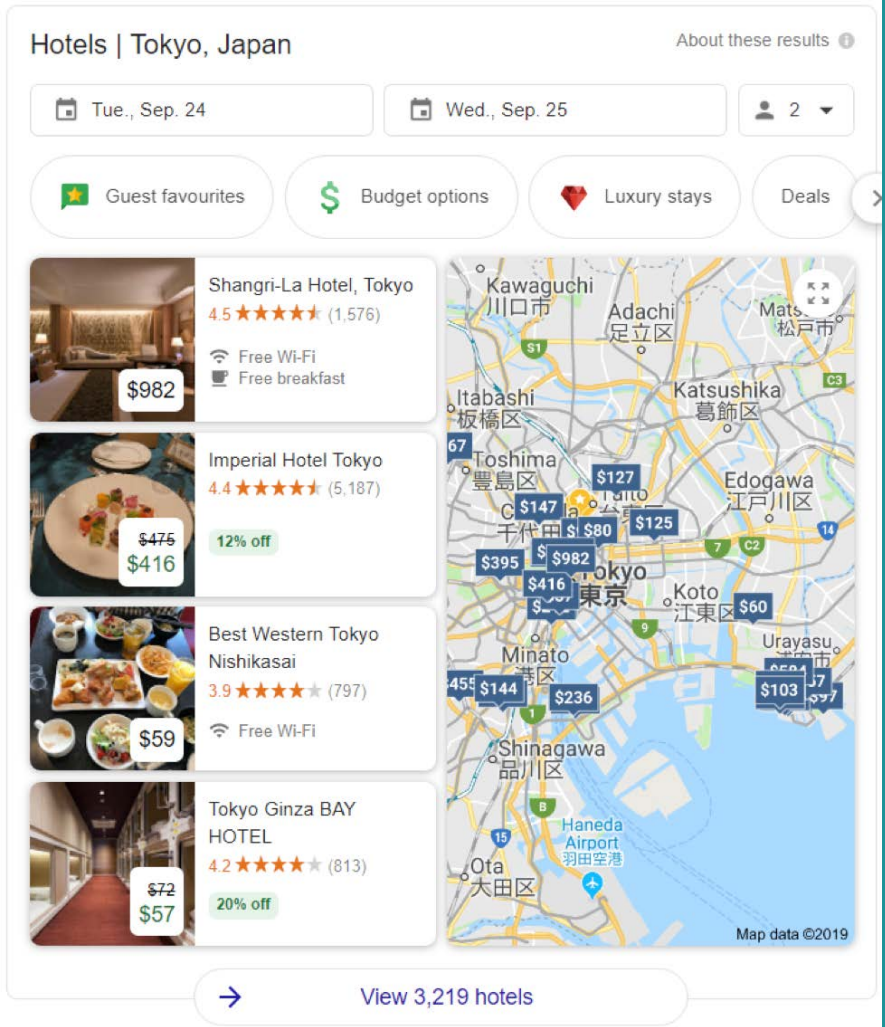
Google has added events into the featured snippet area we discussed above as Section A. This just happened last February though it was on mobile prior to that.

So ... get your event schema up-to-date.

And if we're going to the Cherry Blossom Festival in Tokyo we probably need a place to stay.

TRAVEL

If you run hotels or are just looking for a place, a quick query on Google and you'll find in the layout:



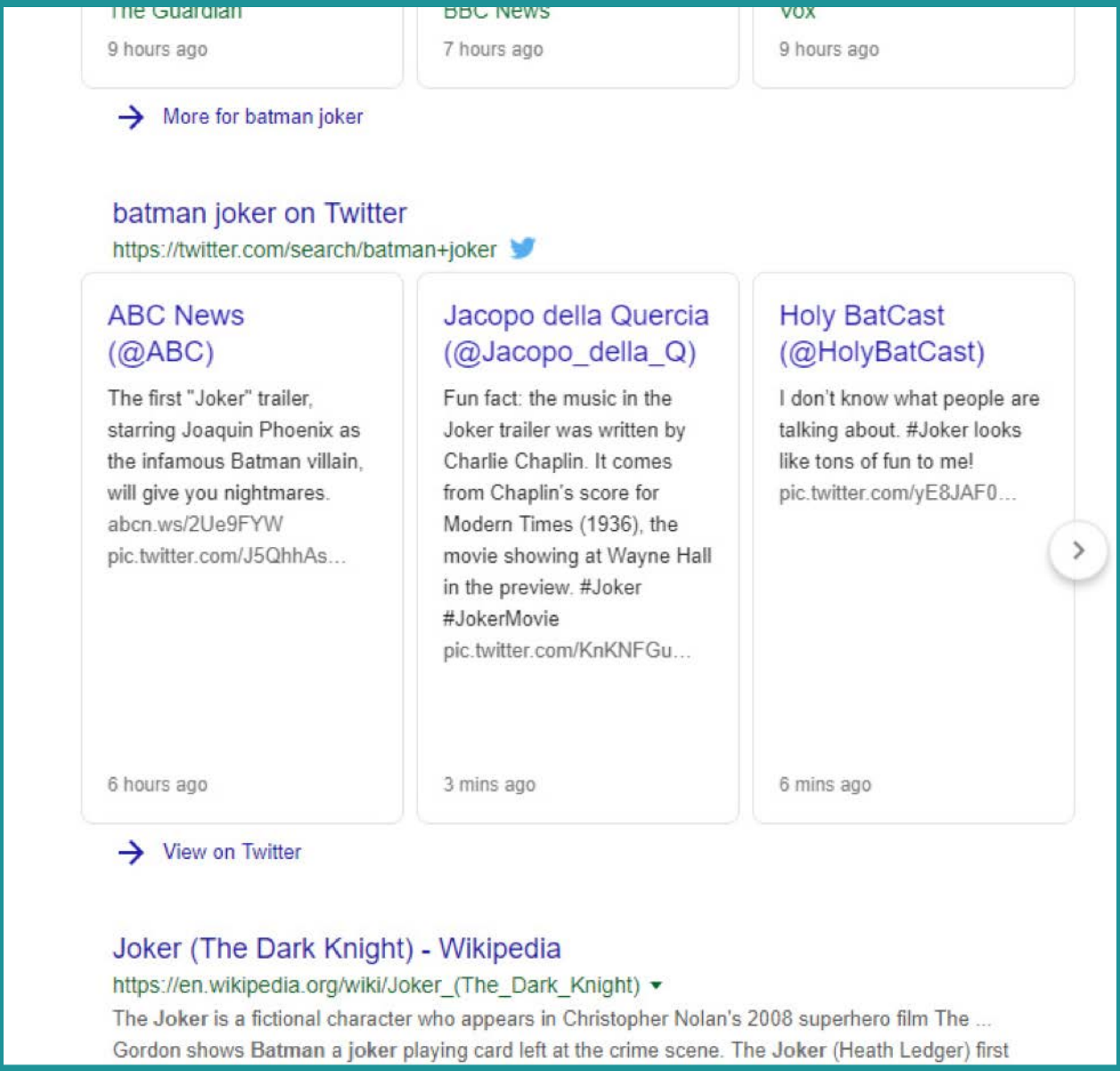
A carousel and map lend the familiar options and you're guided down the path towards a conversion.

While this is similar to the traditional map layout, the volume of filters and options make it a massive threat to those in the travel sector.

The way into this section is paid via **Google Hotel Ads.**

TWITTER

For topics that are trending we see:



Where Google is pulling in tweets from fairly strong Twitter accounts right into the search results.

AND MORE...

As noted above, I know I'm likely missing many.

In future pieces, I'll be diving into some specifics on news, maps, images, and video but if you can think of any content blocks or zones I left out... please don't wait until then.

We'd love to see them posted on our Facebook post on just this subject, which we've set up **here.**

WHY DOES THIS MATTER?

You may be wondering why it matters. You're focused on the top 10 organic links or maybe the featured snippets so why does any of the rest concern you?

The first and most obvious answer is that knowing the various zones and elements on the page informs you as to the opportunities there. In fact, for the first query I entered above there are many opportunities buried in there.

Think about the query and the layout and question always whether there are elements on the page that would steer the users to subsets.

I asked, “what is the civil war”. Might I be sidetracked by a “People also ask”?

Could I get pulled into YouTube? What suggested searches might I click as Google tries to keep me from journeying to page 2?

In these are hidden opportunities.

But there’s more than that.

Within many of these sections, you’re being told specifically how Google is connecting the dots on your topic.

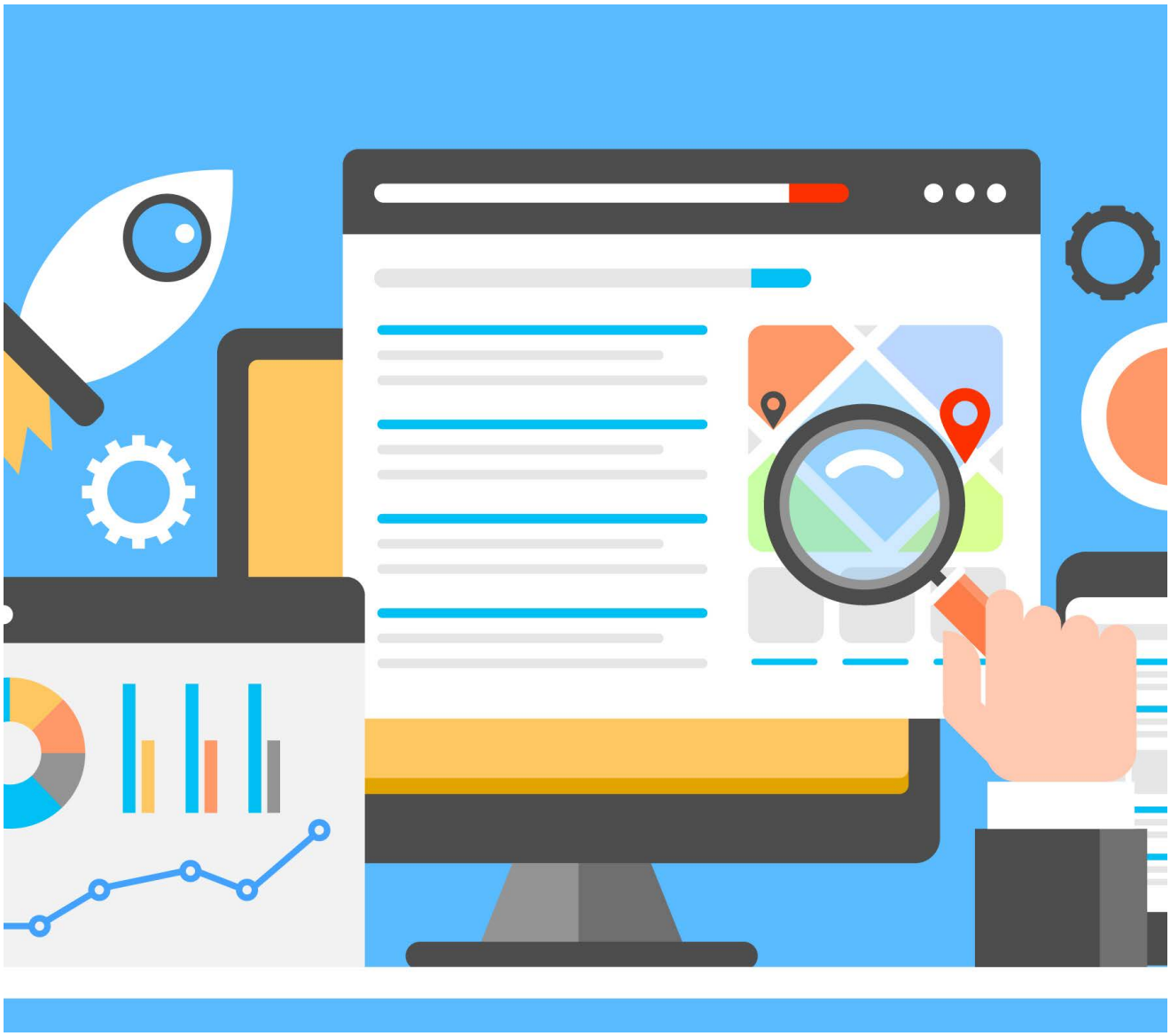
For broad topics think of what the “Searches related to” (G) section is telling you. Think about what the Related Entities (F) mean and how they relate to the content you should be including on your site.

For narrower topics think about what the “People also ask” (C) and Knowledge Panels (B) are signaling.

If people are “also asking” questions that Google has deemed relevant to the questions you ask, should you not be answering them too?

Do the “Related Searches” (K) not tell you what entities Google considers related? Heck, they say so right in the naming of the section.

And of course, look to the formats. If Google wants to provide results in specific formats for specific queries, it’s likely that the searchers and responding to them. That means they’ll respond to you if you produce it.



Looking at the SERPs can tell you a LOT about how Google is connecting entities together and if they are, then doing the same can't help but send a strong signal of relevancy.

When thinking about your content strategy... look to the SERPs.

NOT TO MENTION MOBILE SERPS

I've used a lot of examples here and they've all leaned on the desktop. What can I say, I had to choose one and it was easier to get screenshots.

The same basic elements exist on mobile, but you will often find them arranged in a different order.

Pay attention to this of course as it tells you how relevant each zone is on different devices. If you're ranking highly in organic on mobile, you may be buried beneath more videos and carousels than on desktop.

Knowing this will help you understand your traffic and where to put your efforts based on where your market conducts their queries.

What it tells you about your subject however remains constant, however, it may advise you on how that content is formatted.

CH 8

HOW SEARCH ENGINES ANSWER DIRECT ANSWERS WITH 'USEFUL RESPONSES' & RICH RESULTS

The primary goal of a search engine is to help users complete a task (and, of course, to sell advertising).

Sometimes that task can involve acquiring complex information. Sometimes the user simply needs a single answer to a question.

In this chapter, you'll learn how search engines determine which category a query falls into and then how they determine the answer.

HOW SEARCH ENGINES QUALIFY QUERY TYPES

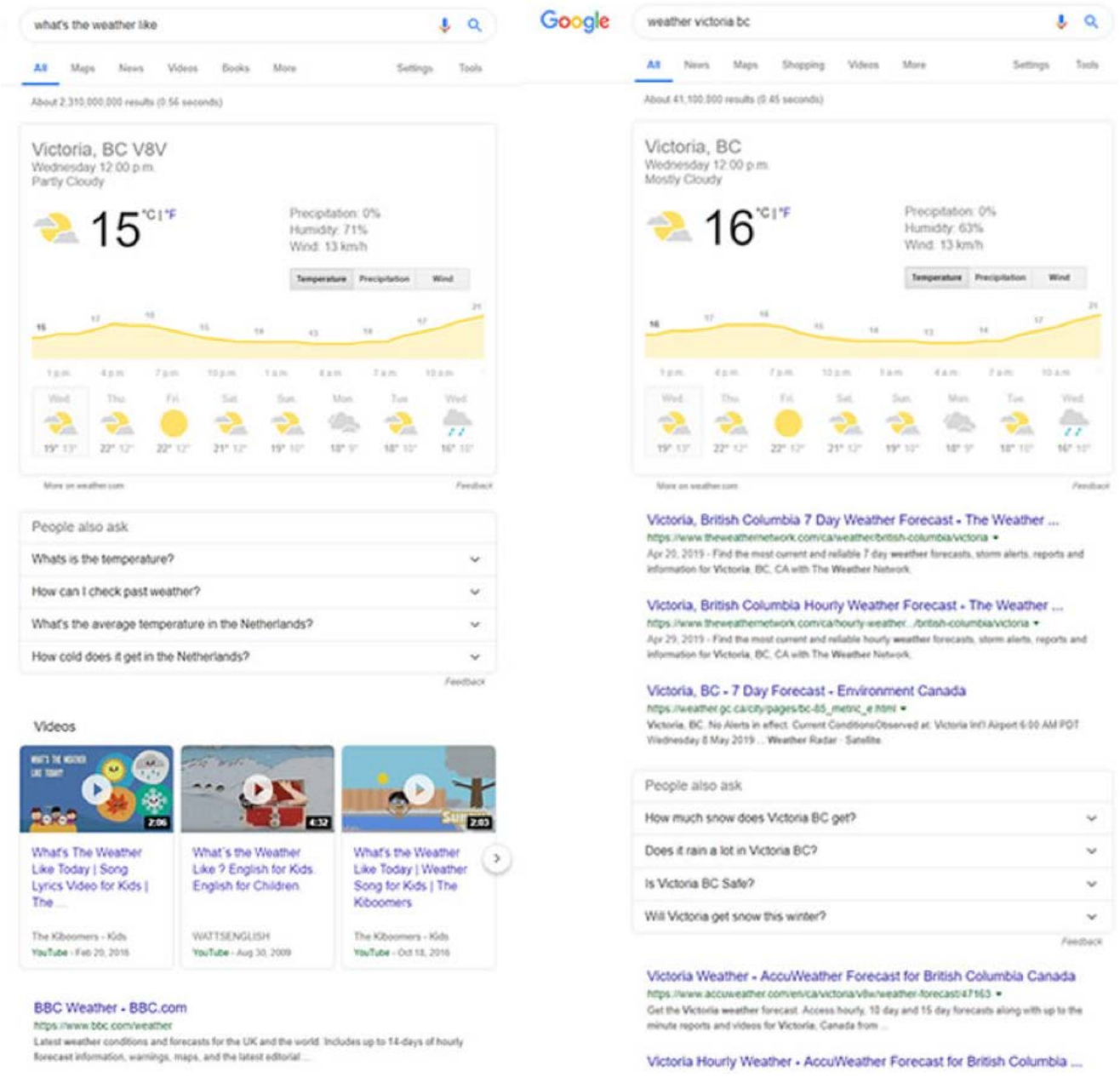
Entire articles, or likely books, could be written on just this question alone. But we're going to try to summarize that all in a few hundred words.

Just to get it out of the way, **RankBrain** has little to no role here.

So what's actually going on?

At its core, the first step in the process is to understand what information is being requested. That is, classifying the query as a who, what, where, when, why or how query.

This classification can take place regardless of whether those specific words are included in the query as illustrated by:



So, what we see happening here is two things:

• Google has determined the user is looking for an answer to a question as the likely primary intent.

• Google has determined that if that is not the primary intent of the user, that the secondary intents are likely different.

You may be wondering how the search engines can determine that the user is asking a question in the second example above. It isn't built into the query, after all.

And in the first example, how do they infer that the user is looking for information on the weather in their location as opposed to just in general.

There are a number of systems that connect and provide data to create this environment. At its core, it relies on the following:

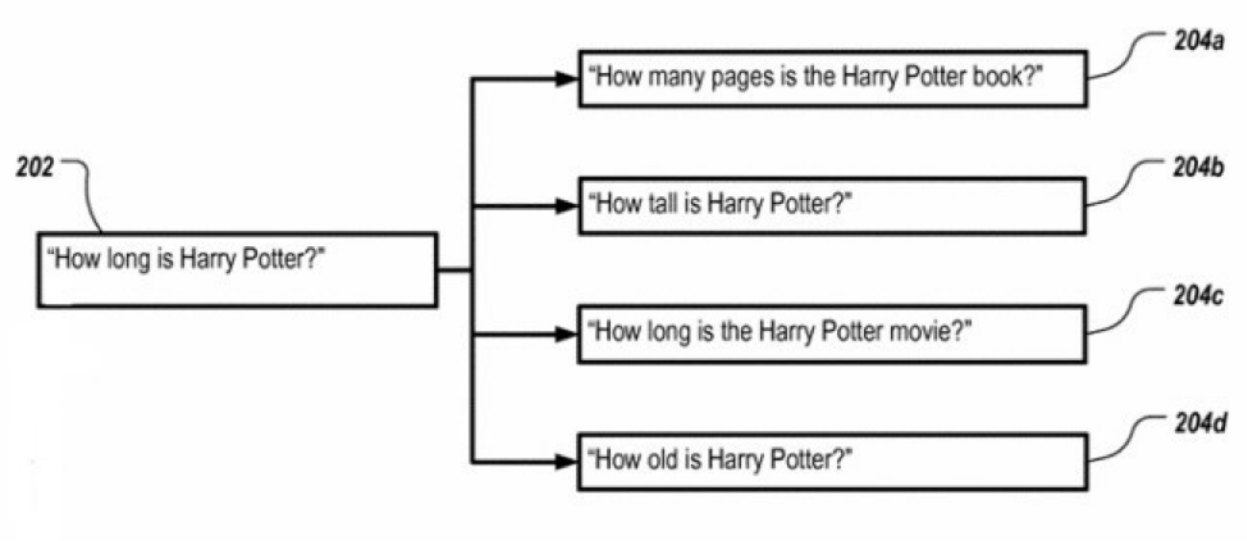
CANONICAL QUERIES

We tend to think of a query as a single request with a single response. This is not the case.

When a query is run, if there is not a known-good likely intent or when the engine may want to test their assumptions, one of the methods they have at their disposal is the creation of canonical queries.

Google outlined the process in a patent granted in 2016 titled, **"Evaluating Semantic Interpretations Of A Search Query"** (link is to my analysis for easier reading).

In short, the problem is summarized in the following image:

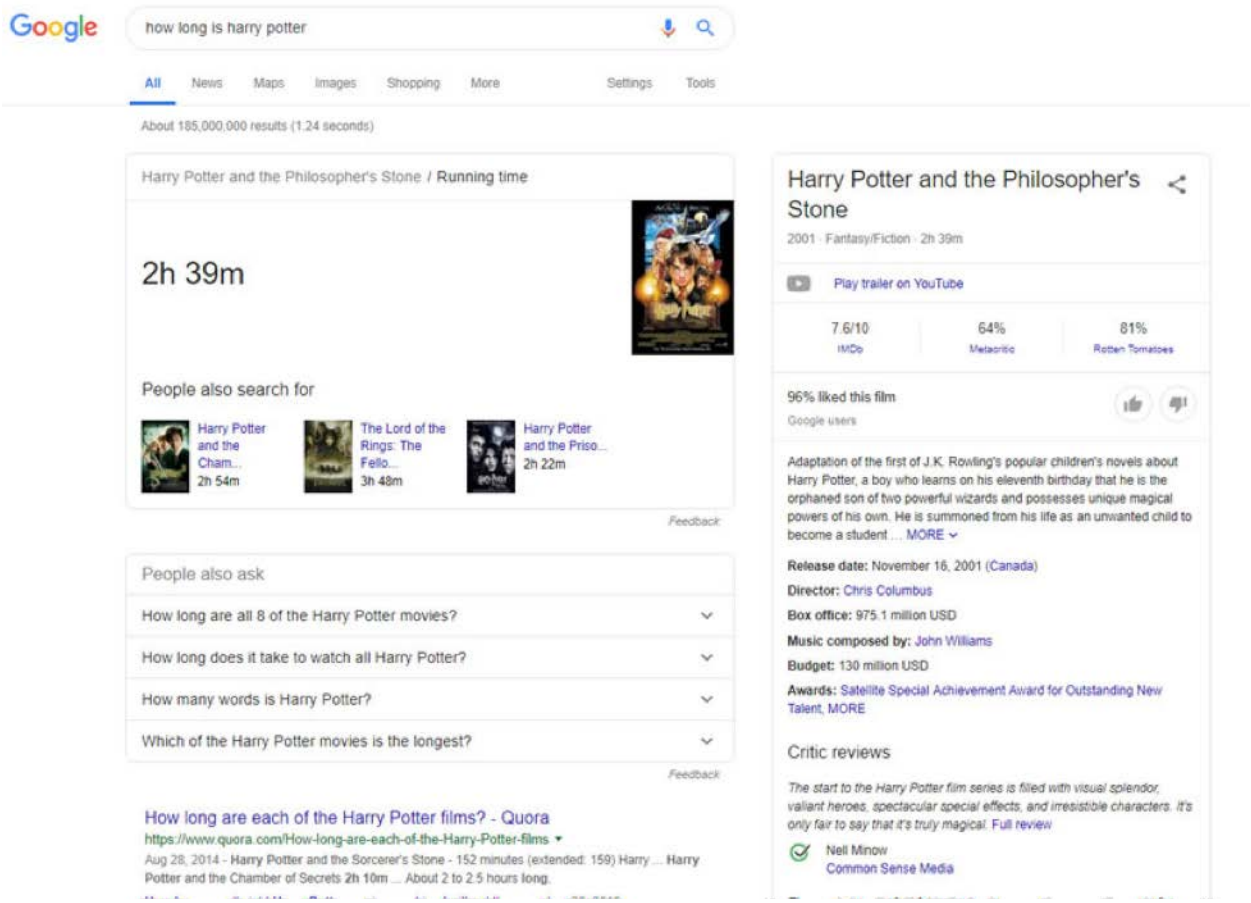


One query with multiple possible meanings.

In the patent, they outline a process by which all possible interpretations could be used to produce a result. In short, they would produce a result set for all five queries.

They would compare the results from the queries 204a, 204b, 204c and 204d with the results from 202. The one from the 204-series that most closely matches that from 202 would be considered the likely intent.

Judging from the current results, it seems 204c won:



Which would have required two rounds of this process. The first to select movies, the second to select which movie.

And the less people that click on a search result from this page, the more successful the result would be considered which is outlined in the patent in the statement:

"USING SEARCH RESULTS TO EVALUATE THE DIFFERENT SEMANTIC INTERPRETATIONS, OTHER DATA SOURCES SUCH AS CLICK-THROUGH DATA, USER-SPECIFIC DATA, AND OTHERS THAT ARE UTILIZED WHEN PRODUCING THE SEARCH RESULTS ARE TAKEN INTO ACCOUNT WITHOUT THE NEED TO PERFORM ADDITIONAL ANALYSIS."

Relative to the context of the patent, this is not saying CTR is a direct metric. In fact, this statement is more akin to what John Mueller meant when answered to a question about Google using user metrics:

"... THAT'S SOMETHING WE LOOK AT ACROSS MILLIONS OF DIFFERENT QUERIES, AND MILLIONS OF DIFFERENT PAGES, AND KIND OF SEE IN GENERAL IS THIS ALGORITHM GOING THE RIGHT WAY OR IS THIS ALGORITHM GOING IN THE RIGHT WAY."

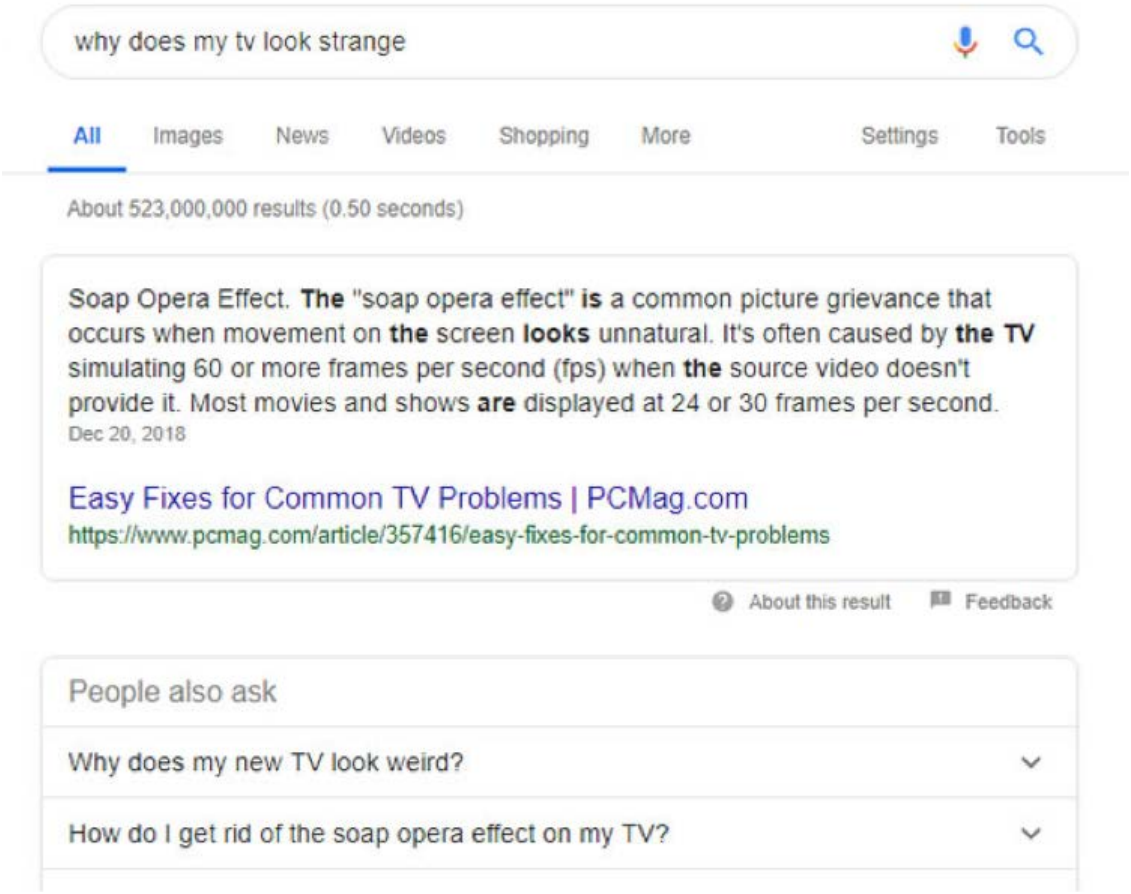
Basically, they don't use it to just the success of a single result, they use them to judge the success of the SERPs (including layout) as a whole.

NEURAL MATCHING

Google uses neural matching to essentially determine synonyms.

Basically, neural matching is an AI-driven process that allows Google (in this case) to understand synonyms from a very high level.

To use their example, it allows Google to produce results like:



You can see that the query is for an answer to why my TV looks strange which the system recognized as a reference to the "soap opera effect." The ranking page doesn't contain the word "strange."

So much for keyword density.

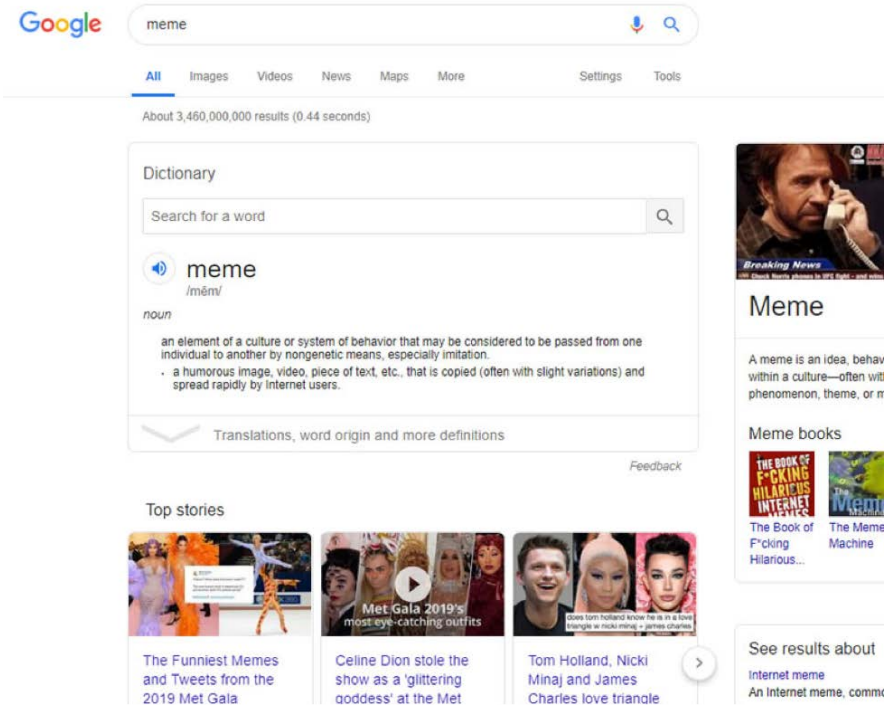
Their AI systems are looking for synonyms at a very complex level to understand what information will address an intent, even when it's not specifically requested.

SITUATIONAL SIMILARITIES

There is a variety of examples and areas where situational context comes into play but at its core, we need to think of how query intent varies by situational conditions.

Above we mentioned a patent about systems that create canonical queries. Included in that patent is the idea of creating a template. A template that could be used for other similar queries to start the process faster.

So, if it took resources to determine that when someone enters a single word that tends to have a broad context they likely want a definition, they can apply that more universally producing results like:



And from there begin looking for patterns of exceptions, like food.

And speaking of food, it serves as a great example supporting my belief (and I think logic) that it's also very likely the engines use search volumes.

If more people search for restaurants than recipes for a term like "pizza", I believe it's safe to say they would use that as a metric and know if a food product doesn't follow that pattern, then the template may not apply.

SEED SETS

Building on templates, I believe it is very likely, if not certain, that seed sets of data are used.

Scenarios where the engines train systems based on real-world understanding of what people want, programmed by engineers, and templates are generated.

Dave sat down at the Googleplex, wanted some pizza, Googled [pizza], got a top 10 list, thought, "that's silly," and started working with the team on a template.

I haven't actually read anything about seed sets in this context, but it makes sense and most certainly exists.

PAST INTERACTIONS

The search engines will test whether their understanding of an intent is correct by placing a result within an applicable layout and seeing what users do.

In our context above, if a possible intent of the query "what's the weather like" is that I'm looking for an answer to a question, they will test that assumption.

It seems that on a large scale, it's an answer people want.

SO, WHAT DOES THIS HAVE TO DO WITH ANSWERING QUESTIONS?

Great question.

To understand how Google answers questions we needed to first understand how they can pull together the data to understand whether a query is a question.

Sure, it's easy when it's a who, what, where, when, why or how query. But we need to think about how they know that a query like "weather" or "meme" is a query for a specific piece of information.

It is a Five Ws query without any Ws (or an H for that matter).

Once that is established using an interconnectedness of the techniques discussed above combined (and I'm sure a few I've missed), all that's left is to find the answer.

So a user has entered a single word and the engine has jumped through its many hoops to establish that it is likely a request for a specific answer. They are now left to determine what that answer is.

For that, I'd recommend you start by reading **what John Mueller has to say about featured snippets** and work your way forward as applicable to your business.

CH 9

HOW UNIVERSAL SEARCH WORKS

Let's begin by answering the obvious question:

WHAT IS UNIVERSAL SEARCH?

There are a few definitions for universal search on the web, but I prefer hearing it from the horse's mouth on things like this.

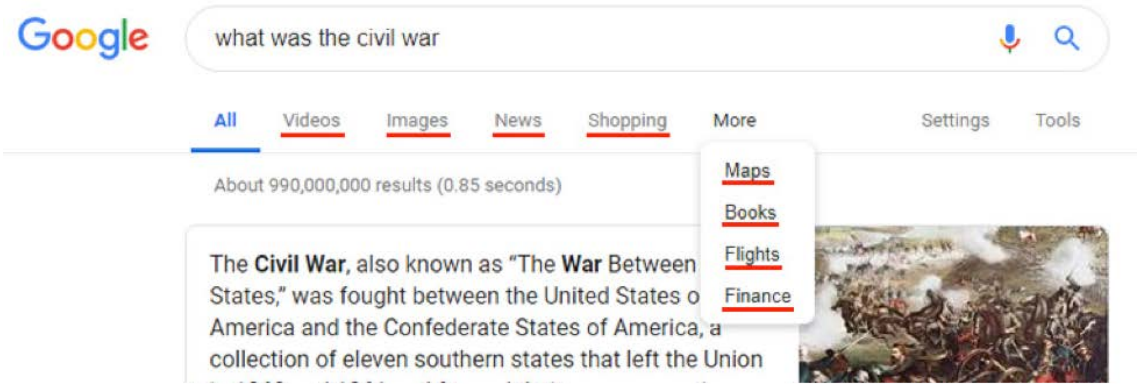
While Google hasn't given a strict definition that I know of as to what universal search is from an SEO standpoint, they have used the following definition in their **Search Appliance documentation:**

> **"UNIVERSAL SEARCH IS THE ABILITY TO SEARCH ALL CONTENT IN AN ENTERPRISE THROUGH A SINGLE SEARCH BOX. ALTHOUGH CONTENT SOURCES MIGHT RESIDE IN DIFFERENT LOCATIONS, SUCH AS ON A CORPORATE NETWORK, ON A DESKTOP, OR ON THE WORLD WIDE WEB, THEY APPEAR IN A SINGLE, INTEGRATED SET OF SEARCH RESULTS."**

Adapted for SEO and traditional search, we could easily turn it into:

"UNIVERSAL SEARCH IS THE ABILITY TO SEARCH ALL CONTENT ACROSS MULTIPLE DATABASES THROUGH A SINGLE SEARCH BOX. ALTHOUGH CONTENT SOURCES MIGHT RESIDE IN DIFFERENT LOCATIONS, SUCH AS A DIFFERENT INDEX FOR SPECIFIC TYPES OR FORMATS OF CONTENT, THEY APPEAR IN A SINGLE, INTEGRATED SET OF SEARCH RESULTS."

What other databases are we talking about? Basically:



On top of this, there are additional databases that information is drawn from (hotels, sports scores, calculators, weather, etc.) and additional databases with user-generated information to consider.

These range from reviews to related searches to traffic patterns to previous queries and device preferences.

WHY UNIVERSAL SEARCH?

I remember a time, many years ago, when there were 10 blue links...

It was a crazy time of discovery. Discovering all the sites that didn't meet your intent or your desired format, that is.

And then came Universal Search. It was **announced in May of 2007** (by Marissa Mayer, if that gives it context) and rolled out just a couple months after they **expanded on the personalization of results.**

The two were connected and not just by being announced by the same person. They were connected in illustrating their continued push towards Google's mission statement:

"OUR MISSION IS TO ORGANIZE THE WORLD'S INFORMATION AND MAKE IT UNIVERSALLY ACCESSIBLE AND USEFUL."

Think about those 10 blue links and what they offered. Certainly, they offered a scope of information not accessible at any point in time prior, but they also offered a problematic depth of uncertainty.

Black hats aside (and there were a lot more of them then), you clicked a link in hopes that you understood what was on the other side of that click and we wrote titles and descriptions that hopefully fully described what we had to offer.

Search was a painful process, we just didn't know it because it was better than anything we'd had prior.

CHAPTER 9

ENTER UNIVERSAL SEARCH

Then there was Universal Search. Suddenly the guesswork was reduced.

Before we continue, let's watch **a few minutes of a video** put out by Google shortly after Universal Search launched.

The video starts at the point where they're describing what they were seeing in the eye tracking of search results and illustrates what universal search looked like at the time.

OK – notwithstanding that this was a core Google video, discussing a major new Google feature and it has (at the time of writing) 4,277 views and two peculiar comments – this is an excellent look at the "why" of Universal Search as well as an understanding of what it was at the time, and how much and how little it's changed.

HOW DOES IT PRESENT ITSELF?

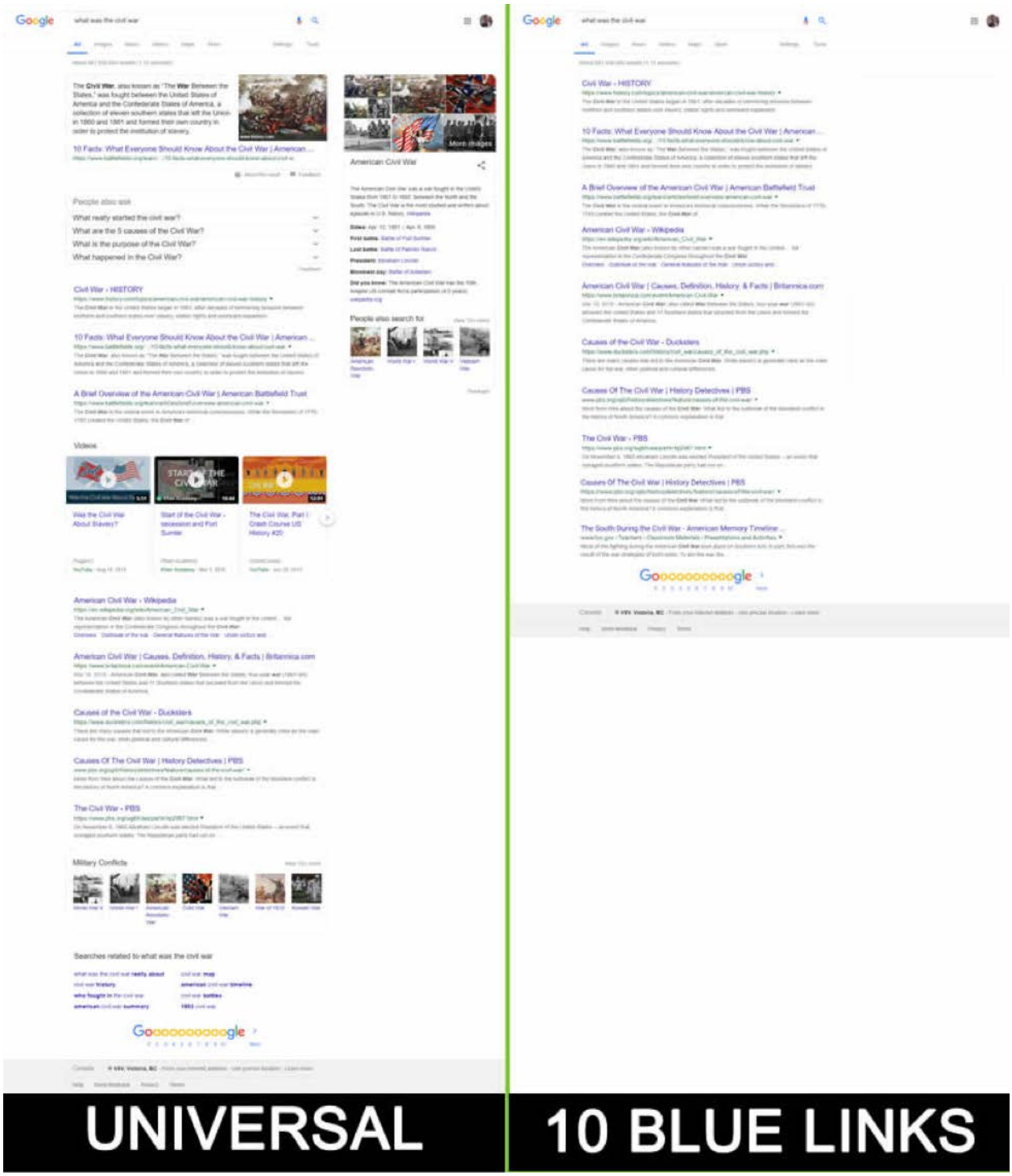
We saw a lot of examples of Universal Search in my in **Chapter 7: How Search Engines Display Search Results.**

Where then we focused on the layout itself and where each section comes from, here we're discussing more the why and how of it.

At a root level and as we've all seen, Universal Search presents itself as sections on a webpage that stand apart from the 10 blue links. They are often, but not always, organically generated (though I suspect they are always organically driven).

This is to say, whether a content block exists would be handled on the organic search side, whereas what's contained in that content block may-or-may-not include ads.

So, let's compare then versus now, ignoring cosmetic changes and just looking at what the same result would look like with and without Universal Search by today's SERP standards.



This answers two questions in a single image.

It answers the key question of this section, “How does Universal Search present itself?”

This image also does a great job of answering the question, “Why?”

Imagine the various motivations I might have to enter the query [what was the civil war]. I may be:

• A high school student doing an essay.

• Someone who simply is not familiar with the historic event.

• Looking for information on the war itself or my query may be part of a larger dive into civil wars across nations or wars in general.

• Someone who prefers articles.

• Someone who prefers videos.

• Just writing an unrelated SEO article and need a good example.

The possibilities are virtually endless.

If you look at the version on the right, which link would you click?

How about if you prefer video results?

The decision you make will take you longer than it likely does with Universal Search options. And that’s the point.

The Universal Search structure makes decision making faster across a variety of intents, while still leaving the blue links (though not always 10 anymore) available for those looking for pages on the subject.

In fact, even if what you're looking for exists in an article, the simple presence of Universal Search results will help filter out the results you don't want and leaves SEO pros and website owners free to focus our articles to ranking in the traditional search results and other types and formats in appropriate sections.

HOW DOES GOOGLE PICK THE SECTIONS?

Let me begin this section by stating very clearly – this is a best guess.

As we're all aware, Google's systems are incredibly complicated. There may be more pieces than I am aware of, obviously.

There are two core areas I can think of that they would use for these adjustments.

USERS

Now, before you say, "But Google says they don't use user metrics to adjust search results!" let's consider the specific wording that Google's John Mueller used when responding to a question on **user signals:**

"... THAT'S SOMETHING WE LOOK AT ACROSS MILLIONS OF DIFFERENT QUERIES, AND MILLIONS OF DIFFERENT PAGES, AND KIND OF SEE IN GENERAL IS THIS ALGORITHM GOING THE RIGHT WAY OR IS THIS ALGORITHM GOING IN THE RIGHT WAY.
BUT FOR INDIVIDUAL PAGES, I DON'T THINK THAT'S SOMETHING WORTH FOCUSING ON AT ALL."

So, they do use the data. They use it on their end, but not to rank individual pages.

What you can take this as it relates to Universal Search is that Google will test different blocks of data for different types of queries to determine how users interact with them. It is very likely that Bing does something similar.

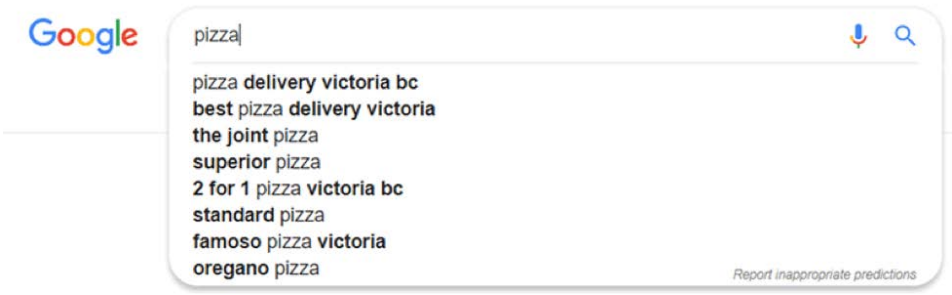
Most certainly they pick locations for possible placements, limitations on the number of different result types/ databases, and have determined starting points (think: templates for specific query types) for their processes and then simply let machine learning take over running slight variances or testing layouts on pages generated for unknown queries, or queries where new associations may be attained.

For example, a spike in a query that ties to a sudden rise in new stories related to the query could trigger the news carousel being inserted into the search results, provided that past similar instances produced a positive engagement signal and it would remain as long as user engagement indicated it.

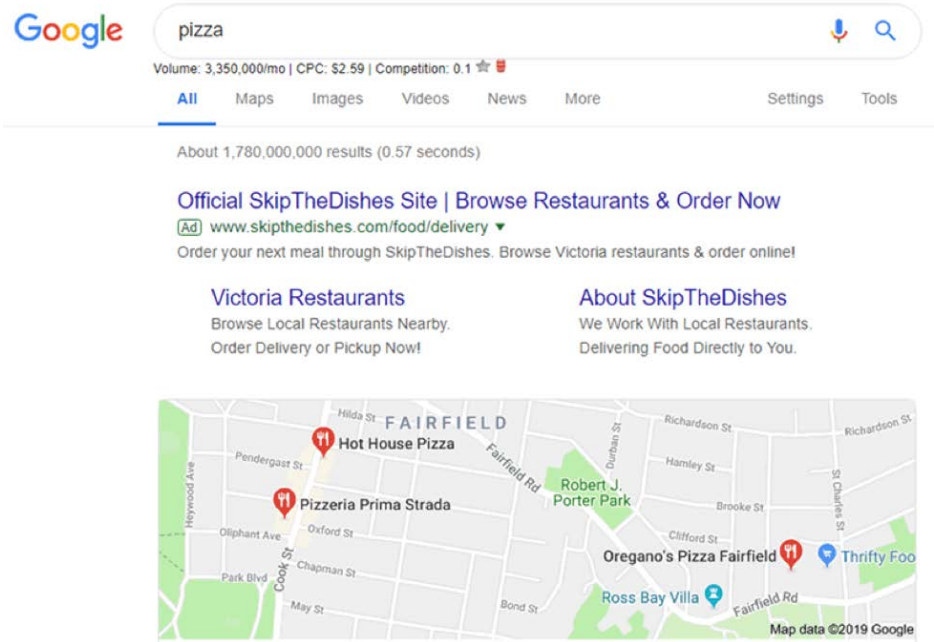
QUERY DATA

It is virtually a given that a search engine would use its own query data to determine which sections to insert into the SERPs.

If a query like [pizza] has suggested queries like:



Implying that most such searchers are looking for restaurants it makes sense that in a Universal Search structure, the first organic result would not be a blue link but:



It is very much worth remembering that the goal of a search engine is to provide a single location where a user can access everything they are looking for.

At times this puts them in direct competition with themselves in some ways. Not that I think they mind losing traffic to another of their own properties.

Let's take YouTube for example. Google's systems will understand not just which YouTube videos are most popular but also which are watched through, when people eject, skip or close out, etc.

They can use this not just to understand which videos are likely to resonate on Google.com but also understand more deeply what **supplemental content** people are interested in when they search for more general queries.

I may search for [civil war], but that doesn't mean I'm not also interested in the Battle at Antietam specifically.

So, I would suggest that the impact of these other databases does not simply impact the layouts as illustrated in Universal Search but that these databases themselves can and likely are being used to connect topics and information together and thus impacting the core search rankings themselves.

TAKEAWAY

So, what does this all mean for you?

For one, you can use the machine learning systems of the search engines to assist in your content development strategies.

Sections you see appearing in Universal Search tell us a lot about the types and formats of content that users expect or engage with.

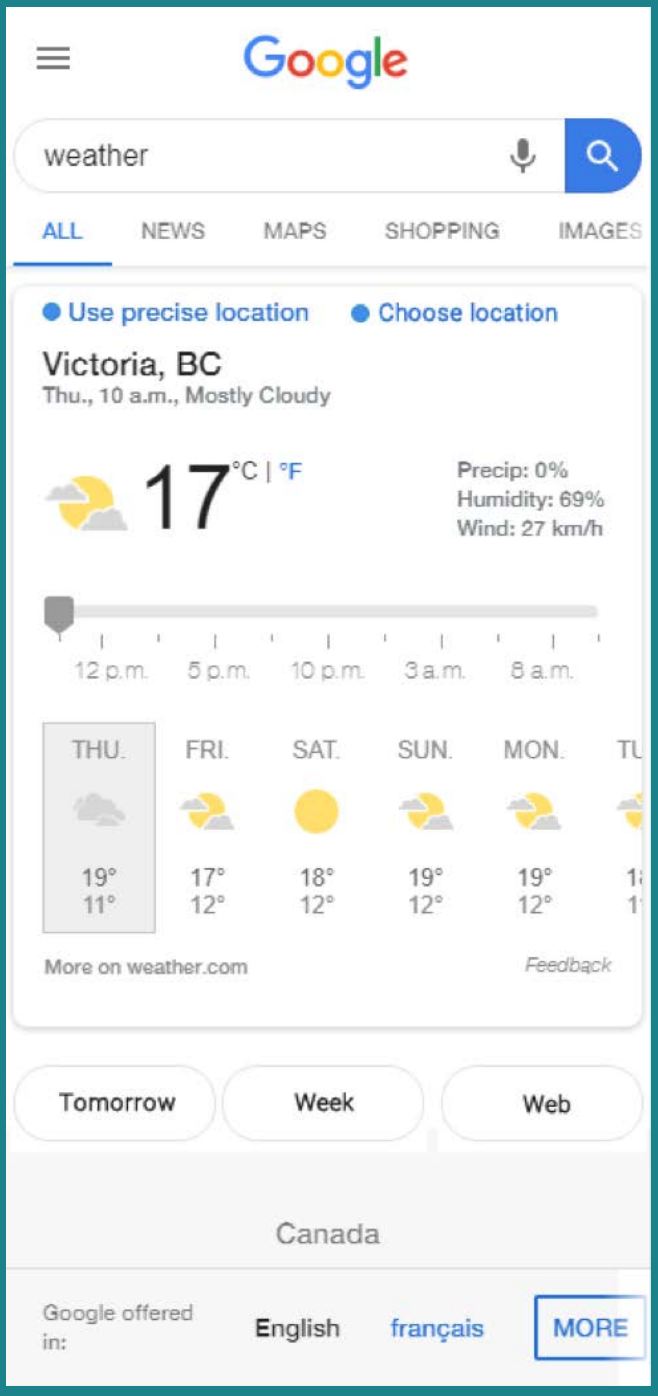
Also important is that devices and technology are changing rapidly. I suspect the idea of Universal Search is about to go through a dramatic transformation.

This is due in part to **voice search**, but I suspect it will have more to do with the push by Google to provide a solution rather than options.

A few well-placed filters could provide refinement that produces only a single result and many of these filters could be automatically applied based on known user preferences.

I'm not sure we'll get to a single result in the next two to three years but I do suspect that we will see it for some queries and where the device lends itself to it.

If I query "weather" why would the results page not look like:



In my eyes, this is the future of Universal Search.

Or, as I like to call it, search.