The Internet and the Search for Truth

By

Mark McGee

I've been providing content for Internet sites for almost 30 years. That includes content for television stations, businesses, non-profits, and personal websites. I know the benefit the Internet can have for people. I also know the downside of Internet usage. That's what I want to address in this series of articles.

Has the Internet helped or hurt journalism? Do we use it at our own peril? How can we be sure the information we discover through Internet searches is true?

Journalism, at its core, is a **search for truth**. The search for truth as a journalist in the 1960s, 70s, 80s, and early part of the 90s was done the 'old fashioned' way. We made scores of phone calls, talked to a lot of people in person, visited many at offices and houses, met informants on back roads and in alleyways, dug through scores of boxes in the back of government storage rooms, and read hundreds of pages of documents for news stories. Then the Internet came along.

Even though the Internet was pretty basic when I started working with it in the mid-90s, I could see the advantages of searching for information online. Government agencies, including local, state, and federal, added more and more information to the Internet. That made it easier and faster to find documents important to a news story or investigation through search engines. The Internet still didn't replace the hard work of good journalism I used years before, but it did become another important tool in the journalist's tool belt.

Before I answer some questions about whether the Internet has helped or hurt journalism, let's take a quick look at Internet history for some clues.

I used *Yahoo! Search* a lot in the 1990s, along with several other search engines. Searching for information back then worked pretty well and it just kept getting better as more search engines came along. Remember these names?

- Altavista
- AllTheWeb
- Archie
- Ask Jeeves
- Excite
- Google
- Gopher
- HotBot
- InfoSeek

- JumpStation
- LookSmart
- Lycos
- MSN Search
- Overture
- WebCrawler
- World Wide Web Worm

I tried most of them and used the ones that seemed best for my searches. As time went on, I found myself using Google more often than the others. So did millions of other people, including journalists. The phrase '“Google It” became synonymous with searching for something on the Internet.

Google continues to dwarf other search engines in market share and volume of queries. However, I highly recommend that journalists use multiple search engines as part of their search for truth. Why? Because of how search engines work.

Some of the earliest search engines were based on people adding information manually. Another system is known as *crawlers*. "Web crawlers" automatically *crawl* the web for content that is new or updated. I remember learning about "spiders" and "robots" as part of the world of providing content for Internet searches. It's one thing to provide content, it's

another for people to find that content when they search the Internet for key words or phrases.

Enter *Search Engine Optimization* (SEO). That's how website owners, designers, and content providers increase the potential that the information on their website pages will rank higher in searches. They use meta tags and other types of coding to "optimize" their pages for search results.

Here's an example from several search engines for how many search results can be found using one word. I searched for the word "truth" to see what the engines found.

- Google - 2,210,000,000 results found in half-a-second
- Yahoo - 182,000,000 results found in half-a-second
- Bing - 101,000,000 results found in half-a-second
- Google Scholar - 3,730,000 results found in half-a-second
- Internet Archive - 213,131 results found in half-a-second

I used about 20 different search engines for this test, but most did not list the number of results found. They do work pretty well, so it's good to list them here.

- AOL
- Ask
- Blackle
- Boardreader
- Brave
- Dogpile

- DuckDuckGo
- Ecosia
- Gibiru
- Gigablast
- OneSearch
- Neeva
- Startpage
- Swisscows
- Webopedia
- YouTube

Using multiple key words in a search is usually the best way to find information for a news story. An Internet “search engine” is basically a software system that searches for information that’s “find-able.”

The information you find in a search may or may not be true. That often surprises people, but not every web page contains truth. Many pages are filled with lies and half-truths. That’s where journalists have to use other skills to determine what information to trust for their stories.

The vast majority of people who search online rely on just the first page of results in a web search, and they often use only one search engine. Most never visit page two or three. Keep in mind that each *results* page contains an average of 10-12 website pages. Using our examples above, clicking on the websites from just the first page of search results means that hundreds, thousands, or even millions of web pages will never be seen.

That's not a good idea for journalists because the information they need to find the truth may be dozens, hundreds or thousands of results pages into the search. There has to be a better way to find truth on the Internet.

Remember the basics of real journalism?

- Curiosity
- Skepticism
- Accuracy
- Objectivity

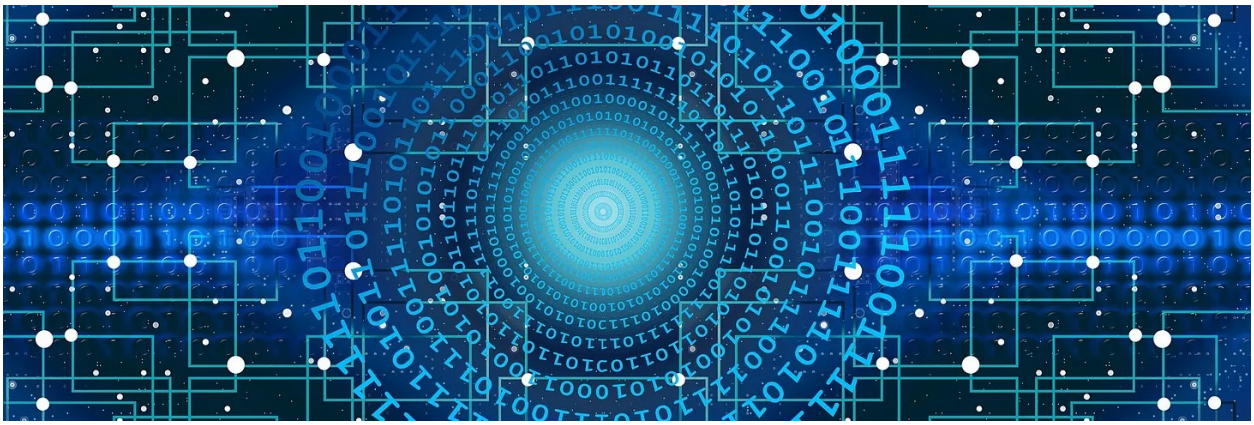
Search engine results often differ from one engine to another. Why? That's both a curious and skeptical question. I'm *curious* about why top search results (those found on the first results page) are different from one search engine to another to another. I'm also *skeptical* about what's behind the way results are *ranked* by different search engines. Search ranking is simply the order in which web pages appear in search results. Many pages are never found in searches because they are buried so deep in the results inquiry.

Algorithms

It is important to understand how *algorithms* affect Internet searches. Remember the “web crawlers?” They travel through the Internet looking at readable pages on websites. The *algorithm* is what creates an index from what the crawlers find. The index that the algorithm creates is what you and I see when we search for one or more key words.

Algorithms can and do more than just create indexes. They can also “learn” about users. How do they do that? Through *cookies*. Cookies are very small files that websites place on your browser. These small files track and monitor where you go on the Internet. Cookies can also “see” what you click on as you visit various websites. The cookies are supposed to improve your online “experience” by remembering your interests and looking for items or stories to show that might like to see. The coded algorithms have the ability to “interpret” what you’re watching or reading and find ways of making more of that content available to you.

Many people love what algorithms do for them. They say it makes their online searching better. Is that true? Is it true for journalists?



Algorithms are explained as “a procedure to solve a particular problem in a finite number of steps for a finite-sized input” (Algorithms Design Techniques).

I first came across the history of an *algorithm* while doing research for a book on another topic almost 20 years ago. The word *algorism* was used by a Persian mathematician in the 9th century AD. He was explaining how Arabic numerals work.

What about Internet *search* algorithms?

> A *search algorithm* … is a procedure that determines what kind of information is retrieved from a large mass of data. *Merriam-Webster*

> The purpose of search algorithms is to check or retrieve an element from a data structure where it's stored. It's that simple. This type of algorithm searches for a target (called the key) in the search space – which can be anything from a shortlist of numbers to a massive database with customer data. *Codete*

Another definition caught my attention because of its simplicity:

> An "algorithm" is nothing more than a set of instructions, just like a recipe or how-to book. 11 Essential Algorithms That Make The Internet Work

I am not a computer scientist, algorithm scientist or engineer. However, I am an investigative journalist and researcher. Which means I learn about a subject by interviewing experts in that field, and doing extensive research. Here are some of my findings from a variety of sources. This is a short list, so you can do more research on the subject if you're interested. I'll list some sources at the end of this article that you may find helpful.

- **Routing** - "is the process of finding a path in a network to send traffic. A number of routing algorithms are out there with names like fuzzy routing, heuristic routing, and adaptive routing, but each one helps make network communication possible." *Business Insider*
- **Entities** - Information that is singular, unique, well-defined, and distinguishable
- **Links** - a connection between entities
- **Filtering** - a type of content control
- **Ranking** - evaluating content according to specific criteria (e.g. relevance, quality)
- **Encryption** - "An *encryption algorithm* is a set of rules by which information or messages are encoded so that

unauthorized persons cannot read them." (*Merriam-Webster*)

Who Are They?

A lot of computer coding is involved in building algorithms, which means humans do the building. Who are they? Here are some descriptions that may help.

An algorithm specialist is a computer scientist who researches and designs algorithms for academic and real-world applications. What Does an Algorithm Specialist Do?

An algorithm engineer is responsible for improving AI applications, to help clients or employers identify patterns or problems in data sets. One of several high-profile artificial intelligence jobs, the role of an algorithm engineer commonly includes the creation, installation, and analysis of algorithms for evaluation purposes. How to Become an Algorithm Engineer

Algorithm Engineers (AE) make a decent salary. I searched for AE jobs and found that the pay range is from about $90,000-$150,000 a year, usually with full company benefits. I worked with an RSO at a high-tech university for many years and met some students who were studying to meet the qualifications for AE. Many of them already had good-paying jobs with high-tech firms before they graduated.

Should Journalists Trust Them?

Journalists should be skeptical about everyone. That's part of our training. It doesn't mean we won't trust anyone. It just means that we need evidence for trust. "Trust, but verify." So, how can a journalist verify whether the algorithms developed for any given website or social media platform are set up for fairness? Remember: Truth is the goal of journalism. What's the goal of people who design algorithms?

Interestingly, fairness is an issue in algorithm development. I came across an interesting article written for Stanford University Engineering. Notice the admission made in the opening to the article.

> Algorithms inform the news you read, the TV shows you watch, and the advertisements that appear on your internet searches – and they also have a say in who gets a bank loan, what medical procedures are covered by insurance, and who gets selected for a job interview.

As algorithms are used to make these decisions, how do we make sure they're fair? And what does fairness even mean? How to design algorithms with fairness in mind

Algorithm fairness (AF) is an actual "thing" in the world of machine learning (AI).

Algorithm Bias - algorithmic bias describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others. Also, occurs when an algorithm produces results that are systemically prejudiced due to erroneous assumptions in the machine learning process. Florida State University Libraries

Here are some insights from a “data scientist.”

Algorithm fairness is the field of research aimed at understanding and correcting biases like these. It is at the intersection of machine learning and ethics. Specifically, the field includes:

- Researching the causes of bias in data and algorithms
- Defining and applying measurements of fairness
- Developing data collection and modelling methodologies aimed at creating fair algorithms
- Providing advice to governments/corporates on how to regulate machine learning

It is also important to understand that approaches to fairness are not only quantitative. This is because the reasons for unfairness go beyond data and algorithms. The research will also involve understanding and addressing the root cause of unfairness. What is Algorithm Fairness?

Here is another interesting admission about fairness, or lack of fairness, in machine learning and algorithm development.

> Our world today is becoming more and more automated. Machine learning, the method of creating computer algorithms that improve (and learn) by experience and data usage, is integrated into our everyday lives to facilitate things like job application screenings and university admissions. There is an increasing need to make sure this data science, including the tools and systems we use, is ethical and fair.
>
> When machine learning isn't fair, the outcome can be detrimental to users and the community. For example, algorithms on social media sites may have sparked political tensions due to skewed or siloed news feeds (and fake news), when the intention was to deliver personalized recommendations for users. What Is Machine Learning Fairness?

Design and Development Bias

What I want you as a journalist to see is that computer and data scientists admit that bias can find its way into algorithm design and development. Where does this bias come from? The same place bias can find its way into journalism — people with biases.

One of the things I've harped on in *Real Journalism* for the last year is the importance of objectivity in journalism. The worst side of journalism is when journalists are biased toward a particular view of what's objectively true. I believe the same is true for people who design algorithms for the Internet and social media.

> Whatever the reason, biases do exist in humans and now they are also passed into the artificial intelligence systems created by humans. 5 Algorithms that Demonstrate Artificial Intelligence Bias

Internet Search Cautions

One of the reasons I encourage journalists to be cautious about believing what they find in Internet searches or on social media is because of how easy it is for people behind the scenes to put their technical “foot” on the scale. Rewriting code to make changes in outcomes is not hard to do, if you know how to do it.

This cuts both ways. Political, racial, and economic biases affect all political parties, races, and socio-economic factors. If journalists are able to do the job of being objective in covering every story, they should be able to recognize when Internet searches or social media posts and comments are biased and possibly inaccurate. However, if journalists are not able to be objective with every story, the potential for falling for biased search results could be a deep pit into which journalists will stumble and fall.

What harm does that do? It harms the journalist’s reputation, the reputation of journalism, and the news consumer who depends on an accurate and objective news media. We can and must do better.

Online Sources

These are some sources you may find helpful in your research about Internet and Social Media algorithms. There's a lot of information available to read, so look at this as a starting point.

Business Insider

Codete

Digital Marketing Institute

Digital Trends

Dynamic Yield

Geeks For Geeks

Ivey Business Journal

Pew Research

Search Engine Journal

SEO Quake

Thales

Turing

UX Pickle

Journalism is hard work

Real journalism is hard work. Hard work. Lots of early mornings and late nights. Lots of weekends investigating leads. Hours and hours going over documents, listening to interviews, talking with witnesses and informants, going over notes time and again trying to make sure that everything is gathered and confirmed. Making sure to report the truth. Hard work. Important work. Sometimes, dangerous work.

Real journalism is not surfing the Internet, checking your social media accounts, getting a few facts, making a couple of phone calls, then writing the story and calling it a day. That's easy. Real journalism is not easy.

Don't get me wrong, please. I like some aspects of the Internet. I've been developing content for television stations, businesses, non-profits, and personal websites for almost 30 years, so I understand the value of the Internet. What I do believe is wrong is how journalists look at the Internet. Let me explain.

Journalism Before the Internet

1983

I was a journalist for almost 30 years before getting involved in the Internet in 1995. Many in my generation have retired or moved on to different vocations where we can still use our skill set. Younger journalists have taken our place. That's the way of life. However, technology is moving at such a rapid pace that many of today's journalists have never known a

world where computers and the Internet did not exist. They grew up surfing the web, playing video games, texting their friends, and taking selfies.

I'm concerned that their life experience may not have prepared them for the hard fact that people in power lie and deceive. I'm concerned that younger journalists may be too reliant on getting their "truth" from the Internet, from social media, and from friends and influencers. That's why I started this journalism blog a year ago. I guess it's one way for an older journalist to help younger journalists.

Today's Gen Y and Z journalists will take journalism in one or more directions. My generation of journalists fought the battles of the past. Today's young journalists are fighting new battles that will take them into the future. My hope is that they will fight those battles for the same reason I fought mine - **to uncover the truth and make it known.**

I say uncover because many people in power cover up what's true. As far back in history as we can go, we find powerful people who lie and deceive. We also find in history a small number of brave people who fought for the truth. The United States of America stands out in history because of its commitment to both the freedom of speech and freedom of the press. The founders of our country wanted people to speak freely, and the press to report the truth.

The Internet can be a great tool for a journalist. It can also be a great deceiver. Some of what I read and watch on the Internet is not accurate or objective. Journalists need to be aware of that and do everything they can to not be duped into reporting something that isn't true. That's one reason reporting news today is so hard. There are so many lies that we have to dig through to find the truth.

Why bother? Because the public deserves to know the truth. They may not like the truth, but they need the truth. They may not want to hear the truth, but journalists must tell them the truth. That's a tough job — hard work that will keep you up at night.

I recently wrote an article about the future of news for Gen Y and Z. The majority get their news from the Internet, especially from social media. That is also true of Gen Y and Z journalists. I get it. I depended on radio, television, newspapers, and magazines for news when I grew up in the 50s and early 60s. It makes sense that you use and trust what is familiar to you. The question for young journalists is whether the Internet and social media are trust-worthy.

I had to ask the same thing about radio, television, newspapers, and magazines when I became a journalist. I worked for some tough news directors in my early days as a reporter and they put the fear of getting a story wrong into our minds and hearts. The worst thing possible was to be called in to the news director's or editor's office because something was wrong in your story. If you were lucky enough not to be suspended or fired, you had to suffer the humiliation of publicly correcting the story and apologizing to your audience or readership.

Believe it or not, that's real journalism. Getting stories right is what we should always do. There is no room for getting a story wrong, then moving along as if it doesn't matter. Unfortunately, too many journalists are getting stories wrong today and proceeding as if being wrong is no big deal. Shame on them. Some of the offenders are veteran journalists. What kind of an example are they giving younger journalists?

Major networks and newspapers have called stories misinformation or even disinformation, only to confirm those same stories as really being true months or years later. No admission of wrong. No apology for publishing lies in earlier stories. No explanation to the public about how or why there were major errors in their reporting. No sense of the damage their social media posts and tweets may have caused. Just a quick note confirming what other media had already reported, then on to the next story as if nothing had happened. That's not *real journalism*.

Proceed with Caution

What all of this means is that journalists should proceed with caution before depending on information they gather from the Internet and social media. Here are some rules that guide me in my research and writing:

1. Use multiple Internet search engines. Dig deep enough into search results to find a variety of views on what you're investigating. Be curious.
2. Use multiple social media sites for research. Question everything that people claim is true. Check, double-check, and triple-check to make sure the people can be trusted and that what they're claiming really did happen the way they say it did. If you have any doubts about something claimed online, talk with people directly. Call people on the phone. Set up Zoom or Skype calls. Visit them in person. Read documents for yourself.
3. Be careful who you trust. Just because someone you like or admire personally makes a truth claim online doesn't make it true. Their article, post or comment may contain some information that's true, but be skewed toward a particular persuasion that's not true. Be skeptical. Ask tough questions. Make truth your goal, not being liked by people you like. Real journalism can be a lonely business. I learned long ago that reporting truth will eventually upset just about everyone you know. If being popular with a group of people is a driving force in

covering stories, then stop and decide what you want to do with your life. Journalism may not be for you.

4. Don't report anything you find online until you've done your own confirmation. Truth claims are not true just because a well-known journalist, celebrity, or friend says it. Be skeptical about everything you read and hear until you confirm it through a journalistic process that can hold up in a court of law. That's where some of your stories may take you some day. You'll be glad you were skeptical and cautious when you face a judge to answer questions about how you gathered information, how you confirmed it, and how you reported it. Also be ready to go to jail to protect a source. I speak from experience. When a judge looks you in the eye and asks you for the second time to give up a source or be in contempt of court and go to jail — you know being a journalist is a serious profession.

1960s Radio News, © Mark McGee