# Article Title: Google’s Search Box Changed the Meaning of Information

# Article Date and Author: Aug 2nd 2023, Elan Ullendorff

# Article Content:

The hallway is bathed in harsh white, a figment of LEDs. Along the walls, doors recede endlessly into the distance. Each flaunts a crown of blue light at its base, except for the doors you’ve walked through before, which instead emit a deep purple. But these are but specks of sand in the desert of gateways.

You are searching for something.

You prepare yourself for an arduous journey. Before the first door you come upon a pedestal. The box that lies on the pedestal gives airs of gildedness despite being as plain as the walls that surround it. It isn’t adorned with a title, but its name echoes in your mind, intuitively: the Answer Box. A plaque reads:

I have crawled through each and every door. Not just the doors in this hallway, but the doors in every hallway in existence, the doors within doors, as well as some doors that I dare not show you, doors that would make you flee in terror. I have seen everything. I am impartial. I have your best interests at heart. I understand what it is you want to know and it is knowable. I have the answer that you seek.

Your finger caresses the latch.

Cataloging the web was doomed from the start. In the summer of 1993, Matthew Gray created the World Wide Web Wanderer (WWWW), arguably the first internet bot and web crawler. During its first official attempt to index the web, the Wanderer returned from its expedition with 130 URLs. But even in the baby years of the internet, this list was incomplete.

To understand how a simple web crawler works, imagine making a travel itinerary that contains three cities: New York, Tokyo, Paris. While visiting each destination, listen for any mentions of other places and add those to your itinerary. Your world crawl is complete when you have visited all of the cities on your ever growing list. Will you have seen a lot of places by the end of your journey? Undoubtedly. But will you have seen the whole world? Almost certainly not. There will always be cities, or entire webs of cities, that are effectively invisible to this process.

A web crawler similarly consults a list of URLs and recursively visits any links it sees. But the resulting index should not be confused with a comprehensive directory of the internet, which does not exist.

I have a theory of technology that places every informational product on a spectrum from Physician to Librarian:

The Physician's primary aim is to protect you from context. In diagnosing or treating you, they draw on years of training, research, and personal experience, but rather than presenting that information to you in its raw form, they condense and synthesize. This is for good reason: When you go to a doctor’s office, your primary aim is not to have your curiosity sparked or to dive into primary sources; you want answers, in the form of diagnosis or treatment. The Physician saves you time and shelters you from information that might be misconstrued or unnecessarily anxiety-provoking.

In contrast, the Librarian's primary aim is to point you toward context. In answering your questions, they draw on years of training, research, and personal experience, and they use that to pull you into a conversation with a knowledge system, and with the humans behind that knowledge system. The Librarian may save you time in the short term by getting you to a destination more quickly. But in the long term, their hope is that the destination will reveal itself to be a portal. They find thought enriching, rather than laborious, and understand their expertise to be in wayfinding rather than solutions. Sometimes you ask a Librarian a question and they point you to a book that is an answer to a question you didn't even think to ask. Sometimes you walk over to the stacks to retrieve the book, only for a different book to catch your eye instead. This too is success to the Librarian.

There are book reviews that say "I read this so you don't have to" (Physician), and others that say "I read this and you should too" (Librarian). There are apps that put you in a perpetual state of simmering, unrealized wanderlust from the comfort of your couch (Physician) and others that inspire you to get up and go (Librarian).

A search engine, at its core, is a product that tries to help you visit pages made by humans, quintessentially Librarian. In a 2004 [Playboy interview](https://kottke.org/plus/misc/google-playboy.html), Google cofounder Larry page was unequivocal in his assertion that he wanted to “get you out of Google and to the right place as fast as possible.” But over the past 10 years, let's just say Google has gone to medical school. The answer is king; a mere link is nothing more than failure of technology.

Google Search launched five years after the World Wide Web Wanderer, and its main innovation was its PageRank algorithm, which created a trustworthiness score for each website based on how often other "trustworthy" sites linked to it; this score was used not only to decide which sites to index and how often, but also how highly to rank them in search results.

I'd like to emphasize here the utter audaciousness of this undertaking. I remember when Google [first announced](https://maps.googleblog.com/2007/05/introducing-street-view.html) in 2007 that it would take 3D scans of the world in order to power Google Street View. The task felt impossibly, absurdly immense. But over the course of a decade, whether through sheer economic might, or creative use (or exploitation) of labor, [Google managed to do just that](https://www.blog.google/products/maps/cheers-street-views-10th-birthday/). Or at least, it's convinced us that it has.

Every large-scale archival project is a Shakespearian tragedy that always ends the same way: incomplete. It requires players with the hubris to go on every night, as well as an audience willing to suspend disbelief, to believe in a corporate overlord's omniscience and omnipresence. Because there are more streets than it is realistic to scan. And even once scanned, a street continues to evolve: Buildings are torn down, trees grow taller, empires fall. The signified distances itself from the signifier. So difficult decisions need to be made. And hidden within those decisions are ideologies about which places are worth saving.

The number of websites outnumbers miles of road by many orders of magnitude.

Building an index, while onerous, is only part of the battle. There is also the problem of processing your search query into a list of results. Usually this involves natural language processing (NLP), a set of techniques that help computers interpret human communication. A rudimentary NLP algorithm might split the query "baking a loaf of bread" into individual tokens (baking, a, loaf, of, bread), remove any commonly occurring words that don't add much obvious meaning to the query (baking, loaf, bread), reduce words to their base form to better match word variations (bake, loaf, bread), and expand the query to include common synonyms (bake, cook, prepare, make, craft, loaf, bread).

But the more sophisticated NLP techniques that Google uses today involves wielding a concoction of interconnected machine learning algorithms that predict which results will be most useful to a searcher. The underlying goal is to understand a user's "intent" using any contextual clues at its disposal: current events, and the user's location, search history, language, device. When a user searches for the word "mars," are they searching for information about the planet, the God, the gene, the chocolate bar, the present-tense verb, or the city in Nebraska?

Of course, natural language is a bit of a misnomer. There is nothing "natural" (in the colloquial sense) about the way we talk to Google. We wouldn't walk over to a friend and bark "italian restaurant nearby" or "what watch netflix romcom." [In the words](https://static1.squarespace.com/static/5a6135761f318d1d719bd5d9/t/5b2536342b6a2886441759d5/1529165365116/JOHN_CULKIN.pdf) of the media scholar Father John Culkin, "we shape our tools and, thereafter, our tools shape us.” Put differently, we evolve to ask our questions in ways that we think our machines can answer them, and over time, privilege questions that are technologically solvable. Can Google ever really understand what our intent is? Can we?

A piece of software that interprets your intent and returns a list of links from a large index is a perfectly usable search engine. However, since the early 2010s, Google has embraced a radically different vision of what a search engine can be: one that can respond directly to questions directly on the results page. This feature has been referred to using a slew of confusing, ever-changing names (rich answers, direct answers, instant answers, quick answers, featured snippets, knowledge panel), but for our purposes we’ll use the colloquial umbrella category: the Answer Box.

[The Knowledge Graph](https://support.google.com/knowledgepanel/answer/9787176?hl=en), a semantic network that perceives the world in terms of discrete entities containing structured data, plays a pivotal role in Google's pursuit of this vision. Under the Knowledge Graph, for example, the band Boygenius is associated with genres, record labels, a discography, images, a list of links and videos, and contains the members Julien Baker, Phoebe Bridgers, and Lucy Dacus, who are each themselves considered entities in the Graph with their own associated data.

To cast a slightly wider net of answerable questions, Google also uses a technique it calls [Passage Ranking](https://developers.google.com/search/docs/appearance/ranking-systems-guide#:~:text=Google%20uses%20automated%20ranking%20systems,a%20fraction%20of%20a%20second.), which picks out specific excerpts from pages that might answer a user's question, whether or not it's the focus of the page. Passage Ranking can tell me, among other things, how Boygenius met ("Julien and Lucy performed on the same bill in Washington, DC, followed by Julien meeting Phoebe a month later"), where the band's name came from (“men are taught to be entitled to space … a 'boygenius' is someone who their whole life has been told that their ideas are genius”), and pluck out of a 1400-word [New Yorker profile](https://www.newyorker.com/culture/culture-desk/the-raw-devotion-of-julien-baker) that Julien Baker is “five feet tall and a hundred and five pounds.”

The vision of the world that these rich results represent is one in which everything worth knowing is unambiguous and perfectly atomizable; call it the baseball-card-ification of knowledge. For anything else, well, for that you'll have to scroll a bit. A 2020 investigation by [The Markup](https://themarkup.org/google-the-giant/2020/07/28/google-search-results-prioritize-google-products-over-competitors) found that almost half of Google's mobile results page on the most popular queries was taken up by links to Google's own properties via sections like the “knowledge panel,” “people also ask,” and “featured snippets.”

All of these technologies—web crawling, PageRank, Natural Language Processing, the Knowledge Graph, and Passage Ranking—converge to convince us of a sequence of lies: I have seen everything. I am impartial. I have your best interests at heart. I understand what it is you want to know and it is knowable. I have the answer that you seek.

The Answer Box's decade of glory, at least in its current form, may be coming to an end. Google has [announced](https://blog.google/products/search/generative-ai-search/), to much fanfare, that it is experimenting with injecting generative AI into the results page. This will enable Google to present answers to more oblique queries, like "tell me what makes boygenius' music unique or special," or "write a poem using the titles of unreleased boygenius tracks," queries that we now might associate more with ChatGPT.

Ask ChatGPT a question, and you will be given a convincing-sounding answer, what Neil Gaiman calls "[information-shaped sentences](https://twitter.com/neilhimself/status/1639610373115375616?lang=en)." When I asked it to give me examples of how different cultural and historical contexts shape the definition of creativity, it readily rattled off 10 vague but coherent examples of differing expressions of creativity across time and space. But when I asked it to point me to the source of its knowledge about creativity and Indigenous Australian "Dreamtime" stories, it could only say "as an AI language model, I have been trained on a large dataset of written text, including books, articles, and other documents from a diverse range of fields and sources … I don't have direct access to specific sources that I have been trained on." It then began to list some books I might read, many of which were invented whole cloth. Generative AI is far from the beginning of Google’s foray into Physician-based search, but it just may be the straw that breaks the Librarian’s back.

There is nothing inherently wrong with a Physician. Diving into rabbit holes is time-consuming, and sometimes, with a trusted source, it is worth discarding context to get to the root of understanding. The problem is when that Physician is not a person or a population of people but a monolithic cluster of machine learning algorithms. When we talk about AI, the speed at which we run toward or away from context becomes amplified, and we run along with the three horsemen of generative text—misinformation, economic exploitation, and creative rot—all of which are enlivened by context collapse and allergic to depth.

But even scarier is the soft apocalypse of a truth that's reduced to trivia.

There is the kind of commodifiable Physician-truth you’d get from an encyclopedia entry: Visit five different webpages and they will tell you the same melting point of gold. But there are other kinds of truth as well, the kinds inherent in the poetry—not poems, mind you, but poetry—of everyday context. There is truth in the aesthetic sensibilities of a webpage, in a text’s surroundings, and in a writer's voice. It's the truth of a speaker's involuntary gestures, the twitch of a lip. Truth in the way words feel tossed about the top of your tongue, in the slanting of letterforms, in slips o fthe pen, in (the volume of the words in) parentheses. A sentence fragment that interrupts a rhythm.

A text changes with the knowledge of its provenance. A text changes with the knowledge of how much work was put into it. A reader finds meaning in atmosphere and timbre in the same way that a parent knows whether a baby is crying out of hunger, fear, or exhaustion, or a heart is moved differently by the same song performed in a new key. Like the keen understanding that persists after you awaken from a dream you can't remember, communing with the messy context of human creativity yields a specter that lingers, haunting you with ambiguity and depth.

The specter is what Tim O'Brien called a story-truth that's "truer sometimes than happening-truth"; Audre Lorde called poetry "the way we help give name to the nameless so it can be thought"; and Maggie Nelson (paraphrasing Wittgenstein) called the inexpressible "contained—inexpressibly!—in the expressed."

And this inexpressible, poetic story-truth transcends mere knowledge. It is the foundation of conversation, the exchange of ideas, critical thinking, serendipity, and properly valued labor. These are the particles that coalesce into a community of care that gives a shit about its inhabitants, into an internet that doesn't sacrifice the complex beauty of communication for the fleeting satisfaction of knowing.

There are [hints](https://www.youtube.com/live/cNfINi5CNbY?feature=share&t=2512) that Google may be more interested in providing context than ChatGPT. And AI can certainly, at least in a technical sense, serve as a force in the direction of depth. But Google's business incentives and search history make me skeptical. Dividing an analog world into discrete digital bites of information means that we spend more time with Google's products. It also makes the information easily recyclable for other platforms, like Google’s voice assistants.

In another world, a web crawler can be training wheels for our own crawling, a language processing algorithm can eschew exaction in exchange for the rich stream of consciousness quality of, well, "natural" conversation, and a search engine can withhold the brick wall of a solution and instead present us with doors.

But instead, I worry that the Answer Box is a premonition of where Google wants to go, a future in which we’re hurried toward destinations, journey be damned, and links are only included out of obligation, rather than invitation. I worry that instead of evoking wonder, our tools will treat our wonder as if it’s an ailment. I worry that this will mean not only a Barthesian death of the author, but a death of the humanmade work itself, human language replaced with its simulacrum. I worry that we’re hurtling toward contextual eradication.

Which technological future do we want? One that claims to know all of the answers, or one that encourages us to ask more questions? One that prioritizes output, or accessibility? One that sees people as a dataset to mine and an inefficiency to overcome, or one that sees them as valuable and worthy of attention?

In being given exactly what we’re searching for, will we lose ourselves?