Report on The Web Search Revolution Symposium

Mark Sanderson
RMIT University
Australia
mark.sanderson@rmit.edu.au

Ramon Lobato
RMIT University
Australia

Kieran Hegarty
RMIT University
Australia

Lisa M. Given
RMIT University
Australia

Chirag Shah
University of Washington
USA

Abstract

This report describes a one-day symposium held at RMIT University that celebrated two milestone anniversaries in Web search: the launch in 1993 of JumpStation, the first modern web search engine; and the 25^{th} anniversary since Google was incorporated as a company in 1998. The report describes the keynote presentations from Microsoft Research leader, Susan Dumais and University of Washington professor, Chirag Shah as well as three panel sessions involving leading researchers across Australia in both academia and industry. The symposium explored the impact and future directions of the web search revolution.

Date: 17 August 2023.

Website: https://www.admscentre.org.au/event/web-search-revolution/.

1 Introduction

The world of search has undergone a remarkable transformation, from simple keyword queries to a complex, all-pervasive automated decision-making system. With 2023 marking two important anniversaries: 25 years since Google's founding and 30 years since the release of the first web search engine, JumpStation, it was a timely opportunity to reflect on the profound impact of search on society. The rise in prominence of ChatGPT and other generative AI systems also offered an opportunity to examine the future of information access. The Web Search Revolution: The Past, Present, and Future of Web Search – Google, ChatGPT, Bing, and Beyond, was a one-day interdisciplinary event that delved into the history, evolution, and future of search. Its speakers, presentations, and panel sessions explored the following questions:

- What has been the societal impact of this near-ubiquitous technology?
- What lessons can we learn from the success and pitfalls of introducing search to large-scale platforms?

- How will generative AI shape the future of information access?

The symposium hosted 21 speakers across two keynotes, three panel discussion sessions, and one public debate event. A total of 168 people attended the event with 100 attending in-person and 68 attending online. The Symposium was recorded, its sessions can be watched online.[1] Summaries of two of the keynotes and three of the panels are provided here.

2 Keynote – The History of Web Search and its Impact

Susan Dumais: Technical Fellow & Managing Director, Microsoft Research, USA

To provide some context, as late as the 1980s, most information seeking happened in the physical world, such as browsing through library stacks or printed catalogues. In libraries, books were arranged based on classification systems like the Dewey Decimal or the Library of Congress system. Search for phone numbers might happen in printed directories like the White and Yellow Pages. Online search did exist before the web, systems like Minitel from phone companies provided access to directories and local information online, and services like LexisNexis and Medline provided access to content from some verticals. However, these online systems were not as prevalent or comprehensive as web search engines would become. With the rapid growth of the web starting in 1991, came web directories and web search to help people find all that information. Web search started with JumpStation developed by Jonathan Fletcher at the University of Stirling in Scotland in 1993. It was a crawler-based search engine that actively and automatically gathered content instead of waiting for content to come to it. It used simple natural language interfaces and had a web form for specifying queries. However, JumpStation folded within a year of its inception due to a lack of funding and interest.

Between 1993 and 1998, there was a proliferation of search engines after JumpStation, including WebCrawler, Lycos, Yahoo!Search, AltaVista, AskJeeves and others. Each search engine had its own approach: relevance ranking, term proximity, semantic search, and other techniques were employed by the different engines. The rapid evolution of search algorithms and experiences required an understanding of who is using the systems and what problems they were trying to solve.

[1] https://www.youtube.com/playlist?list=PLE_y90GftjpYZ1q5UMCqhX9Sxid0hM_9_, Accessed: 2023-11-12

An early analysis of web search logs revealed some surprising findings. One analysis conducted by Andre Broder and Craig Silverstein at AltaVista found that queries and sessions were often very short and that queries were not equally likely – a few occurred many times within a day while the majority of queries occurred only once. Another analysis by Broder [2002] showed that not all queries were informational as expected. People were not just searching for in-depth information but also using search engines for purposes including navigation (i.e., getting to web pages) and transaction (i.e., doing things). These findings challenged the assumption that web search was primarily about seeking information.

Although Google was incorporated in 1998, research on it started at Stanford University as a project called BackRub in 1996. Google introduced several innovative approaches to web search including matching on anchor text and the PageRank algorithm, which helped Google provide more relevant and accurate search results compared to other search engines at the time. Google also exploited a successful business model, based on search advertising. Google's focus on – scale, speed, relevance, and simple user experience, which included for the first time, "*a summary for each result that highlights where your query matched*[2]" (i.e. snippets [Tombros and Sanderson, 1998]) – set a new standard for web search. Google gained popularity and quickly grew in prominence as a search engine.

Over the next two decades, web search innovations continued. Context and personalization in search was exploited and a growth in mobile searching started to emerge. Query spell correction, auto-completion and query suggestions became common. Ranking algorithms evolved and increasingly became machine learned eventually incorporating neural methods. The snippets that Google innovated with evolved into answers presented in results leading eventually to a more question answering approach.

One cannot present a reflective talk without also considering the emergence of generative AI and its potential impact on web search. Large language models have broadened people's imaginations about the possibilities of AI and have sparked discussions about their potential applications. Generative AI models, such as GPT4, are one component of larger applications such as web search. These models are being used in applications like ChatGPT, GitHub co-pilot, and Bing chat, among others. There is an on-going challenge of verifying and grounding generated content. Automatically generated content can lack accuracy or reliability. Ensuring the quality and trustworthiness of generated content becomes is crucial in a web search context.

It is becoming clear that bridging the gap between generative AI models and an application like web search requires an orchestration layer that connects the prompt or query with the desired end-user application. In search, an orchestration layer can help ground and verify the generated content, providing context and provenance for the search results.

Considering the future of web search, the integration of digital and physical search – a "*control F for the physical world*" – is yet to be achieved. Also, unifying information across systems would overcome the profound frustration of having information scattered across various apps and platforms. Web search itself will continue to need to meet the changing needs and aspirations of searchers.

[2] http://blogoscoped.com/archive/2006-07-27-n45.html, Accessed: 2023-11-12

3 Panel – The Impact of Search on Society

Ramon Lobato: Associate Investigator ADM+S, RMIT University, Australia

The day's first panel, "The Impact of Search on Society", explored the significance of web search for our institutions, markets, and societies. Bringing together experts in computer science, digital communication, political science and intellectual property law, the discussion provoked a spirited debate and revealed the often very different questions that scholars in these fields bring to the topic of search.

Prof Kimberlee Weatherall (University of Sydney Law School) began by reflecting on the legal decisions that have shaped search businesses over the last three decades, and the importance of understanding the two-way interaction between technological and legal developments. According to Weatherall, everything about search could have been different had the legal environment around search been differently structured; for example, if intellectual property had been more strongly enforced to prevent indexing of open web content, if compulsory licensing had been introduced to require payments to publishers, or if key acquisitions by major tech companies had been blocked under antitrust law. Weatherall also reflected on the legal responsibility of search engines, noting that the permissive environment under which search evolved may be reconsidered by future courts and legislators as the social harms of have become clearer in recent years.

The theme of responsibility was also taken up by Prof Alistair Moffat (computer science, University of Melbourne), who questioned the responsibility of technical experts to tackle social problems and systemic effects. As Moffat jokingly remarked, "I'm a computer scientist. We don't do 'society'." Moffat emphasized that, for much of the history of web search, the role of computer science has been to develop technical solutions and improvements to improve efficiency of search engines. This work is no longer dominant in terms of simple indexing and querying task, but still important in terms of dense and generative models. For Moffat, the question of technology design and its social effects cannot be addressed only by computer science but rather needs to emerge as the result of dialog between multiple disciplines.

Political scientist and communication scholar Louisa Bartolo (Queensland University of Technology) emphasised that the ordering of information is a profoundly political process. Citing foundational work in internet studies by Introna and Nissenbaum [2000] on the politics of search engines, Bartolo explored current questions of governance and regulation in search. Noting that the implicit policy of "leaving search to the market" is now everywhere in question, Bartolo pointed to the new transparency obligations for very large online platforms and search engines built into the EU's new Digital Services Act, which also imposes a legal obligation on those companies to mitigate "systemic risks". For Bartolo, this represents a historical moment in the evolution of search, as policymakers reconsider laissez-faire legal settings under which search evolved.

This theme of power and governance was developed further by Prof Dan Angus (Queensland University of Technology), a computer scientist who works across communication and cultural studies. Angus shared findings from the ADM+S Centre's Search Experience Project, which has examined issues of representation and bias in web search results related to Australian and global politics. Reflecting on these findings, Angus observed that longstanding assumptions that have guided curation of search results for news content may now need to be questioned as we move further into an age of climate crisis, arguing that it may no longer be sufficient to aim for a cross-

section or middle ground of opinion. For Angus, the concept of "we" must always be interrogated and extended to include not only web search users, but also non-users impacted in various ways by the structures of search engines.

A lively Q&A followed the panel discussion. Audience members pressed the panellists on a range of issues, including how to account for the diversity of national governance in search and whether the rise of web search has meant an end to the notion of professional expertise as conceived by the architects of the Enlightenment.

4 Keynote – Generative AI and the Future of Information Access

Chirag Shah: Information and Computer Science, University of Washington, USA

Generative AI tools have started impacting the area of information access in significant ways. Some of these ways are welcome changes that are decades in coming, whereas some others are problematic and even harmful. The question is can we minimize those harms enough and leverage the benefits of generative AI to charge a new era with even wider, deeper, and fairer access to information? To answer this, we need honesty and not hype about these tools and technologies; being able to ask tough questions and be prepared for tough answers. For example, we could ask what assurance we have that an LLM has captured diverse enough, current enough, and comprehensive enough content in a given domain for it to be a reliable and trustworthy source of information. We could and perhaps should ask if an image generation tool is respecting privacy and providing attributions as appropriate. And we could ask if a user, and that may be even us, really understands what a chatbot is capable of and what it's not. Finally, before we get too excited about AGI or even just a smarter AI tool, we should ask why we need it and how we will ensure a level of equity, fairness, and accountability that does not go against our collective societal values. These are just some of those tough questions and we have to understand that the answers many not come easy or positive. But this is one place where "move fast and break" or "ask for forgiveness instead of permission" is not a good idea for many of us in a longer term.

About 20 years ago, movie The Time Machine, based on the classic book by H. G. Wells by the same name, came out. It features a hologram librarian named Vox 114 who is possesses all the world's knowledge and can interact with a user in natural language and gestures. This is one possible vision of our future with information access and AI. Another possibility was envisioned by Swedish philosopher Nick Bostrom around the same time 20 years ago. Bostrom presents a thought experiment in which a paperclip manufacturing AI system ends up destroying the world because that's how it could keep optimizing the production of paperclips. Depending on your inclinations, one or both of these may seem far-fetched, fantastic, or frivolous, but they serve the purpose of showing two extremes that are already taking shape in our psyche today. Many think that the current information access systems are heading toward Vox 114, where an all-knowing AI will be all we will need to find, use, and understand any information we need. But there are also many who believe that the same kind of AI development has the potential to destroy us like that paperclip maximizer. The truth, I believe, is somewhere in the middle. Rather than focusing too

much on either extreme, I suggest we ask a more practical question – how do we ensure that we benefit from AI with all it has to offer while minimize the risks?

There is no one answer for this, but I know we must start somewhere. We can start by understanding our past and assessing our present. We know that self-regulation is not something we have done well, whether it is technology or energy. We understand that technological and societal advancements need to happen with everyone's participation and not through the views of a select few. We can project that the fundamental human values such as safety, privacy, equality, benevolence, and universalism will continue guiding us and they need to be instrumental in any AI that we build. With these in mind, I propose the following three ethos:

1. Conformity: AI must understand and adhere to the accepted human values and norms.
2. Consultation: To resolve or codify any value tensions, AI must consult humans.
3. Collaboration: AI must be in collaboration mode by default and only move to take control with the permission of the stakeholders.

I propose we ensure that any AI that we build is coded with the currently accepted human values and norms (conformity). As the AI starts operating and finds tensions among any values, instead of making decisions on its own, it explicitly consults humans (consultation), thus avoiding any value judgments that go against our desired outcome or learning ill-intended behaviors. Finally, if and when an AI system replaces a human or takes control, it must do so with explicit permission of the stakeholders (users, policymakers, designers as appropriate), and relinquish that control when desired by the same stakeholders.

5 Panel – What did the Web Search Revolution Revolt Against? Exploring the Contested History of Search

Kieran Hegarty: RMIT University, Australia

In a panel titled "What did the Web Search Revolution revolt against?", Mark Sanderson, Julian Thomas, and Lisa M. Given explored how searching for information evolved in light of political, social, and economic changes throughout the second half of the twentieth century. To structure the discussion, the panellists focused on three different times and places that show what search was building on and, in many cases, reacting against.

To illustrate how search had changed over the past seventy years, the panel commenced by screening a clip of a 1956 film made in Sheffield in the United Kingdom, called Books in Hand.[3] Released in 1956 by the Sheffield City Council Libraries Committee, the film vividly illustrates the crucial role municipal libraries played in servicing the diverse needs of an emerging information society. Books in Hand shows that while many of the information needs of post-war society mirror contemporary needs, the systems and supports used to service them have changed dramatically. The film illustrates how searching for information was free for library patrons but involved manual, labour-intensive, catalogue-driven processes, and librarian support. The questions we now expect to solve in minutes through a web search took hours, days, or even weeks to answer, and required specialist searching expertise by trained library professionals.

[3] https://www.yfanefa.com/record/2585, Accessed: 2023-11-12

This led to a discussion of how advances in computing and a changing economic and political environment altered who serviced information needs and how. While libraries continued to provide important information services, the second half of the twentieth century also saw a rise in subscription search systems provided by private corporations. This was linked to a decline in funding for public institutions and the rise of powerful markets for the storage and provision of information. Dialog, developed at Lockheed Martin in the 1960s and now owned by Clarivate, remains one of the best-known examples of these systems.[4]

Many of these pre-web information services were so complex that only trained specialists could search, and consumers often paid for the results. The panel illustrated the technical complexity of these systems through the example of The Financial Times' FT PROFILE. The system allowed users to search articles in every UK broadsheet newspaper over a 5-year period, which involved a complex process of remembering typed commands to select a collection and using individual words to reduce the list of documents returned. Articles were presented in date order, leaving readers to scan for the most relevant items. As well its technical complexity, the costs of these services were prohibitive. While FT PROFILE made valuable information rapidly accessible to people outside establishment business circles, the cost was exorbitant. In 1990, access cost £1.60 a minute or £4.65 ($5.60 USD) at 2023 prices [Poynder, 1992]. In this context, the emergence of free (albeit advertising-driven) search services that emerged in the web era can be seen as a reaction against the expensive (and thus exclusive) information services that emerged throughout the second half of the twentieth century.

The emergence and growing popularity of the World Wide Web in the 1990s again changed search. The rapid proliferation of information on the web called for systems to help users search through this glut of online information. The panellists discussed how decades of information retrieval research laid the groundwork for the emergence of search engines, including the 1994 book Managing Gigabytes, which underpinned the principles of web search [Witten et al., 1994]. The book states that: "*the end result of applying the techniques described here is a computer system that can store millions of documents and retrieve the documents that contain any given combination of keywords in a matter of seconds or fractions of a second.*"

Indeed, Managing Gigabytes was used by Larry Page and Sergey Brin to design Google's iconic architecture. Page and Brin's successful application of the lessons from decades of information retrieval research shows that web search didn't emerge from nowhere. For decades, researchers had imagined a search that was fast, sorted documents by relevance, and was accessible to everyone.

After launching in September 1998, the Google revolution was in motion. People were attracted to the simplicity of the search box, as well as a novel presentation of results that summarised how the retrieved pages matched the query [Tombros and Sanderson, 1998]. Its simplicity can be seen as a reaction against the crowded design of many of its competitors, which featured colourful banner advertisements, news headlines, and weather predictions. In terms of its functionality, Google Search stood out from a crowded search market for several reasons. Anchor text allowed search queries to not only be matched with the text within a page but also with other text linking to that page and PageRank ranked results by counting web links in a page. These were significant advances, and Google's popularity quickly surpassed competitors such as AltaVista and Yahoo

[4] https://clarivate.com/products/dialog-family/, Accessed: 2023-09-29

Search. With more than 90% of the market share today, Google remains the most popular search engine.[5]

With Google's expansion as a market player into a range of services and products, and the emergence of powerful AI tools for fulfilling information needs, the future of search is wide open. This panel on the conditions that led to the ubiquity of web search suggests that the future of search will not simply be decided by developments in technology. The history of search shows that how information is stored, sorted, shared, and accessed is also shaped by shifting information markets and the changing expectations and priorities of societies and governments. With AI search tools on the horizon, we must continually interrogate the tensions between public access and increasingly powerful commercial information markets. Doing so will help us answer what may soon become a pressing question: "what did the AI Search Revolution revolt against?"

6 Panel – Imagine the Year is 2050... How will we Experience Search?

Lisa M. Given: Director, Social Change Enabling Impact Platform, RMIT University, Australia

Our panel engaged in longer-term stargazing, to consider what search might look like in another 25-30 years. While generative artificial intelligence (and such tools as ChatGPT) have captured media attention in recent months, we engaged in an interactive discussion to imagine what the future might hold. Prof Lisa Given started the conversation by showing a clip from Steven Spielberg's (2002) Minority Report. Based on Philip K. Dick's 1956 science fiction novella, the movie is set in Washington, DC, USA in 2054, where the "Precrime" unit's chief investigator (played by Tom Cruise) uses a transparent touchscreen to search through criminal files and photographs for evidence of potential crimes. The investigator stands in front of the large screen and uses full-body gestures to push and pull the materials from view; he organises the documents into some semblance of order, very much embodying Spielberg's vision of the search activity mirroring the actions of an orchestra conductor.[6]

Today, while touchscreens are commonplace, many other aspects of search represented in Minority Report remain fictional. Yet, Lisa explained that the movie serves as an interesting starting point to look ahead to the future and imagine how we might search – in our homes, our schools, and our work lives. She then asked her fellow panellists to reflect on a series of questions to think ahead to 2050 and present their vision of a possible search future. The experts explored how people might find and share information, how workplaces may change, whether dozens of phone apps will continue to define everyday search practices, and who might own and control these (and other) tools and the data they contain.

Prof Jason Potts presented his views through the lens of an economist to consider "how will next generation search create value?" He explained that, fundamentally, an economy is simply a coordination engine for distributed human cooperation, or a matching platform (as that's what a market is, with an economy being many overlapping markets). Jason envisioned the future of

[5] https://gs.statcounter.com/search-engine-market-share, Accessed: 2023-09-29

[6] https://www.salon.com/2002/07/10/underkoffler_belker/, Accessed: 2023-11-12

AI-powered search as a multi-sided market. The current first generation is search as an engine, owned by a regulated company, that enables a human to find information. The next generation is one that partially (or even totally) subsumes the role of economic institutions: finding, matching, contracting, and creating value between more than one human, and possibly many, as well as nonhuman beings, such as smart machines and hyperobjects. He explained that search technology, then, evolves from an extension of the operations of a library and subsumes the matching infrastructure in a web3 based digital economy.

Prof Guido Zuccon's view was that people will not search, in the future, in the ways they search now. Rather than searching for and find documents, we will instead receive information insights. He noted that large language models are already showing us that we can access more than just small snippets of information; we can receive conversations from agents with information included. Guido imagines that agents will anticipate our information needs, but rather than giving us documents (i.e., the types of information we now index in systems), these agents will provide us with packaged services. Agents will interact with APIs and give people complete plans – such as a package of bookings, tourist recommendations, and other information needed for their next holiday. This means people will no longer use search terms and access information sources directly. Rather, the information we need will be mediated by these agents, who will anticipate and fulfill our needs in new ways. This will also raise important questions about who will control these agents and what governance requirements may be needed.

Peter Bailey noted that search remains a difficult problem, because people cannot precisely define what they do not yet know. There are multiple considerations that a retrieval system must satisfy, considering issues such as popularity, authority, and diversity. These are balanced in the face of uncertainty about a person's true intent. Recent developments in AI systems are making it simpler and more fluent or naturalistic to conduct multi-turn interactions between a searcher and the system, which Peter noted is very exciting for what the future might hold. However, he cautioned that doing this remains challenging given the enormous breadth of search needs and ways of expressing them. Peter also explained that, as individuals, we share information about ourselves quite selectively in different social circumstances. And the information seeking ecosystem is distributed across many tasks (and organisations that try to service these tasks), both large and broad or small and narrow; these include general information search vs entertainment vs shopping vs knowledge work and office work vs health vs real estate scenarios, to name just a few. Peter explained that none of these systems will hold or have access to the entirety of our information or interaction history. Only some will be able to afford the high costs of building and optimising truly great search systems. Thus, our retrieval experiences will remain fragmented and of variable quality, although the overall standard of these experiences will continue to improve.

Finally, Dr Johanne Trippas explained that in anticipating the future of search, it is critical to embrace a user-centred approach, placing the user at the core of system development while considering the positive aspects of current systems. In envisioning how we search in 30 years, she believed it is evident that multiple modes of interaction, such as voice, gestures, sensors, image, and video, will converge to provide users with dynamic, personalised, private, and adaptive experiences. Users will seamlessly transition between modalities to cater to their specific needs. She suggested that even though traditional text-based inputs will further evolve into natural language interactions, inherently marked by voice commands, more sophisticated multimodal technologies will surpass these. Search outcomes will manifest in multimodal formats, incorporating text,

images, videos, and interactive elements. Immersive technologies such as virtual reality (VR), augmented reality (AR), and mixed reality (MR) will provide rich and interactive information exploration environments. Johanne explained that while AR and VR technologies are already revolutionising industries, they are poised to exert a more significant influence on search functionalities in the future. Developing these technologies in a user-centred way will empower users to search for information in novel and immersive ways, enabling better access for all.

Acknowledgements

This symposium was supported by the ARC Centre of Excellence for Automated Decision-Making & Society (CE200100005), and funded by the Australian Government through the Australian Research Council.

References

Andrei Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM New York, NY, USA, 2002.

Lucas D Introna and Helen Nissenbaum. Shaping the web: Why the politics of search engines matters. *The information society*, 16(3):169–185, 2000.

Richard Poynder. Online vendors' strategies for reaching the end-user: Ft profile. *Online Review*, 16(3):147–155, 1992.

Anastasios Tombros and Mark Sanderson. Advantages of query biased summaries in information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 2–10, 1998.

Ian H Witten, Alistair Moffat, and Timothy C Bell. *Managing gigabytes: compressing and indexing documents and images*. Van Nostrand Reinhold Company, 1994.