OPINION PAPER

JASIST WILEY

Whose relevance? Web search engines as multisided relevance machines

Olof Sundin[1] | Dirk Lewandowski[2,3] | Jutta Haider[4]

[1]Department of Arts and Cultural Sciences, Pufendorf Institute, Lund University, Lund, Sweden

[2]Department of Information, Hamburg University of Applied Sciences, Hamburg, Germany

[3]Pufendorf Institute, Lund University, Lund, Sweden

[4]Swedish School of Library and Information Science, University of Borås, Borås, Sweden

Correspondence
Olof Sundin, Department of Arts and Cultural Sciences, Pufendorf Institute, Lund University, Box 117, SE-22100 Lund, Sweden.
Email: olof.sundin@kultur.lu.se

Abstract

This opinion piece takes Google's response to the so-called COVID-19 infodemic, as a starting point to argue for the need to consider societal relevance as a complement to other types of relevance. The authors maintain that if information science wants to be a discipline at the forefront of research on *relevance*, search engines, and their use, then the information science research community needs to address itself to the challenges and conditions that commercial search engines create in. The article concludes with a tentative list of related research topics.

"The success of information science, whatever there is, is due to the fact that it did address itself to relevance"

Tefko Saracevic (1975, p. 324)

1 | INTRODUCTION

A Google search for COVID-19 on March 4, 2021, from Lund, Sweden led to a Search Engine Results Page (SERP) that we did not recognize from searches on other similar topics. A large part of the first page seemed to be almost crafted by hand (see Figure 1). The knowledge panels presenting "statistics," "health information," "map of cases," "cases overview," and more information about the disease all contain information aggregated from public health authorities. Only under the heading "top results" are so-called organic results presented. These results are automatically generated from the search engine's database of web content. Yet, we see only one organic result before the list is interrupted by further knowledge panels presenting results from local health authorities or Twitter posts, once again from preselected sources. The list of organic results continues after the knowledge panels (not depicted in Figure 1).

This first cursory inspection of the page already indicates that organic results do not play a particularly prominent role, but that results from sources that in all likelihood are manually preselected dominate the page.

FIGURE 1   SERP for Google search on COVID-19 [Color figure can be viewed at wileyonlinelibrary.com]

When we scroll further down the page, the knowledge panels are gone, and we are left with only organic search results. What is interesting, however, is the fact that even when we consider the organic results alone, no links to "alternative" news or health information providers are included, or for that matter, to any other source that can be assumed to distribute misinformation or disinformation. The organic results consist mainly of newspaper articles, scholarly articles, and warnings from public authorities about the spread of misinformation concerning COVID-19.

We consider the example of the skillfully assembled SERP for COVID-19, and how authoritative sources are prioritized, as a starting point for discussion. This discussion raises a number of central issues regarding web search engines and their role in society. Our intention is to encourage research on web search in the information science community by instigating debate on *societal relevance* (see below) as a complement to other types of relevance. We maintain, in relation to our introductory quote by Saracevic (1975), that if information science wants to be a discipline that is at the forefront of investigating relevance, searching and search engines, more information science researchers need to address the many challenges that commercial search engines impose on society.

2 | WHAT IS IN A SEARCH RESULT?

We share a starting point with many of our scholarly colleagues: online platforms can never be neutral by their very design. A platform would not be a functioning platform if it did not somehow moderate its content or presentation of links (Gillespie, 2018). The inherent need for social media to be moderated (Roberts, 2019), is easy to fathom, but search engines are different. They tend to melt into the background of our practices, and the constitutive role of search engines for society and everyday life is therefore often even more difficult to recognize and challenge. The results an individual gets from a Google search are often considered natural, despite their obvious embeddedness in society's value systems (e.g., Hillis et al., 2012; Lewandowski, 2017; Mager, 2012; Noble, 2018).

All search engines claim to provide relevant results for users. And these users in turn use the search engine in the belief that it will provide relevant results—if they reflect on the issue at all. Yet, who is the user? The point of framing operations such as Google as multisided platforms (Rieder & Sire, 2014) is to highlight the fact that they cater to different user groups or customers, that is, marketers, web searchers, other businesses, a variety of content producers, and so on. All these groups need to be served by the search engine. Thus, their respective interests will also shape what search results are shown on the SERPs (Schultheiß & Lewandowski, 2021) and how relevance is constituted. In light of this, what is *relevance* exactly and how can we understand the relevance of search engine results concerning a public health crisis, such as COVID-19? On Google's own website, relevance is introduced as follows:

> With the amount of information available on the web, finding what you need would be nearly impossible without some help sorting through it. Google ranking systems are designed to do just that: sort through hundreds of billions of webpages in our Search index to find the most relevant, useful results in a fraction of a second, and present them in a way that helps you find what you're looking for. (Google, n.d.)

For many, it might seem self-explanatory what relevance is meant to be here. If we want to know the business hours of a restaurant, the date daylight saving time ends and other undisputed factual information, the question of relevant information is simple. Or so it would seem. But as we mentioned above, Google is a multisided platform that also serves the interests of those producing the content, including business owners, whose livelihood may depend on being visible in search engine results. And this applies to all information providers, whatever their intentions. They are thus coached by Google to produce content that fits the mold provided by the search engine (Google, 2021), and if they do not adhere to the rules, they cannot be found.

That said, even from the perspective of the regular search engine user, many topics are far from undisputed. Not seldom two people can evaluate the same information very differently, depending on what they want to do with the information, their prior understanding of the topic, their ideological position, their age, beliefs, hopes, fears, or even when they encounter the information. In other cases, the topic in question could be disputed in a certain research community or by small fractions in society. Once again turning our attention to COVID-19, a significant number of disputed knowledge claims emerge regarding prevention, containment measures, treatments, vaccine safety, long-term symptoms, origin of the virus, and even the very existence of the disease itself. Some appear outlandish, others are expressions of routine scientific controversy, and others still are rooted in ideological positions or even in the influence operations of foreign powers. How does a search engine such as Google deal with that?

3 | DIFFERENT TYPES OF RELEVANCE

The Merriam-Webster (2021) online dictionary defines relevance as "the ability (as of an information retrieval system) to retrieve material that satisfies the needs of the user." It seems that we have to discuss whose needs we are talking about if we want to understand relevance. But who is in the position to decide if a need is satisfied? To put it bluntly, conspiracy theorists, racists, and terrorists also have information needs. How should those needs be measured? This is a very uncomfortable discussion, but it shows that hiding behind an idea of neutrality is not going to solve the problem. Before going any further, we need to emphasize that the discussion within information science about relevance is advanced and nuanced (e.g., Nolin, 2009; Saracevic, 2016), and it is not our intention here to provide a literature overview. However, earlier literature rarely considers the real-world messiness of commercial, general-purpose search engines and their role in society. This is a huge problem because it means that as a discipline, with our long experience investigating search engines, searching and relevance, we still lack conceptual tools to address the various problematic situations faced by individuals and society as a whole in light of a near search engine monopoly.

Early literature on relevance primarily spoke in terms of *system relevance*, that is, how well a query captures the potentially relevant information in a database. Since the 1990s, the notion of *user relevance* has moved to the foreground (e.g., Hjørland, 2010; Saracevic, 2016). Such a perspective focuses on how individual users evaluate whether the information is relevant in a given situation. The personalization trend in social media and search engines relies on the concept of individual relevance and the opportunity for companies to tailor the feed, and to a certain degree the search results, to fit the calculated interests of individuals, albeit within the bounds of the business model that concurrently caters to advertisers, content providers, and the search engine provider's self-interests. System relevance concerns relevance as an internal quality of the system at hand, while individual relevance concerns the possibility to deliver suggestions based on predictions of information that certain individuals would like to find. The third type of relevance we want to discuss is *societal relevance*, which refers to what is beneficial to society at large (Haider & Sundin, 2019). This comes close to what has been discussed as the "subject knowledge view" of relevance (e.g., Hjørland, 2010; Saracevic, 2016), but it also differs from it. The challenge is when an individual wants to find a specific type of information, for example, conspiracy theories about the vaccine or the virus itself that conflict with the *greater good* of society as a whole. Noble (2018)

discusses this concerning racism and sexism. Continuing the tradition of Hope Olsson, she elucidates how societal values and norms cannot be separated from those of knowledge organization and retrieval systems, such as search engines. You have to consider them together.

The case of the COVID-19 vaccine is an excellent example, as it shows how deeply embedded search engines such as Google and other multisided platforms are in our societies. It also shows that we must continue the work of unpacking our understanding of relevance and move beyond the various models that neatly place different understandings in boxes. We thus argue for the need to explore relevance conflicts or *frictions of relevance* (Haider & Sundin, 2019), not necessarily to solve them but to acknowledge and confront their existence and productiveness.

Throughout the years, Google has had to deal with various instances in which its search results were criticized for advancing racist, sexist, anti-Semitic, or otherwise offensive values. The impression is that Google employs a haphazard whack-a-mole approach.[1] It only reacts in response to media reports and only if these are publicized widely enough to constitute a problem for their brand. The specific issue is addressed—but only after a delay where the problem is explained away and blamed on users. Still, our understanding is that over time, as these cases have accumulated and segments of the public and the media have started to take notice, societal relevance has taken a more prominent role for Google—at least in the West. But how is this done, and what are the implications?

4 | DIFFERENT ANSWERS, DIFFERENT RESULTS

What we have discussed so far raises the question of what strategies a search engine provider may employ to prevent results deemed irrelevant to the greater good of society from showing up in the top results. A first and obvious strategy might be to not include such results in the search engine's database of web pages (the index) at all. However, this would not only raise questions of censorship but would also contradict the claim that the search engine makes all of the web searchable, which underlies every search engine.

Instead, we have to consider strategies that focus on the ranking and presentation of the results. These strategies could follow one or a combination of the following four approaches:

1. Wait for more quality and/or mainstream content to be produced, sometimes directly in response to the appearance of low-quality or disputed results in the search results. While this requires no action from

the search engine provider, it often helps change the tone or ideological direction, mainly in the case of so-called data voids, that is, problematic queries where there are not many documents for a search engine to choose from (see Golebiewski & Boyd, 2018). Given the fact that ranking algorithms consider quality factors to determine the relevance of the results (Lewandowski, 2012), this would lead to higher quality results, thus displacing problematic results from the top positions.

2. Change the ranking algorithm to rely more on "trusted sources." This could mean giving more weight to source credibility, measured, for example, by the quality of inlinks. Algorithm changes like this are quite frequent (see SEO Powersuite, 2020) and affect the organic results overall, meaning the aim of these changes is to improve the overall result quality, not just to demote particular sources.

3. Demoting personalization. As personalization is seen to increase individual relevance at the price of decreasing societal relevance, it could be wise to avoid an excess focus on personalization. Indeed, while the negative effects of personalization have been frequently lamented since the early 2010s (e.g., Pariser, 2011), empirical research has identified the effects of contextualization (e.g., preferring local results), but not many effects from personalization, in Google's top web and news results (Krafft et al., 2019; Nechushtai & Lewis, 2019), although evidence is still too weak to make conclusions.

4. Change the presentation of the results by adding information from trusted sources to the SERP, making organic results disappear from the immediately visible area of the results page. This allows the search engine to have more control over what users will see, as the results in the knowledge panels are not generated algorithmically from an index of web pages but from one or a few manually selected sources only.

In the screenshot shown above (Figure 1), strategies can be seen in interaction: the major part of the SERP is covered by information that is aggregated from trusted sources (in this case, public health authorities) that have been manually selected. When only considering the "area above the fold," that is, the results that can be immediately seen without scrolling, these take up the entire screen. As mentioned above, the organic results are mainly from trusted sources, such as government websites and mainstream news media. This hints that the search engine puts a lot of weight on source credibility and by that, demotes results from less trusted sources. It also seems that Google demoted personalization, as it involves the risk of promoting sources that

are not desirable in terms of societal relevance. While we are very aware that this cannot be proven by considering just one example, evidence strongly suggests that it is a likely scenario. This is as close as we can get when it comes to Google. To maintain some degree of control, as a society and as a discipline, we have to pose some important questions, not least: Who is involved in these decisions, how are they made and how do we know they were made and why (compare, e.g., Mager, 2018)?

5 | IDEAS FOR FUTURE RESEARCH

This short opinion piece points to several important problems and an urgent need for more research. The move by Google towards a greater explicit interest in societal relevance, sometimes at the expense of individual relevance, is an extremely important development for society and one with far-reaching implications. In the case of Google—and in the face of a global pandemic—there are several good reasons for this link between a commercial search engine, public health authorities, and mainstream media. At the same time, the social and political implications of such cooperations are potentially profound and their benevolent character must not be taken for granted. We see an urgent need for the information science research community to engage with these emerging challenges from a variety of perspectives. Possible issues include:

- How can we approach the dilemmas that arise from the normative character of the notion of societal relevance and its implementation in different political and institutional settings? What are the implications for marginalized communities?
- What implications for media and information literacy arise in relation to societal relevance and other societal aspects of search engine use?
- How are web search engines involved in the creation and sustenance of absences and strategic ignorance and how are these implicated in the formation of societal relevance?
- Could the fact that SERPs largely contain results from preselected sources (in knowledge panels or vertical results) hinder users in their information seeking and, in that case, what could the implications be?
- When investigating users in *search sessions*, will users judge the individual relevance of session outcome as better or worse when complex SERPs are generated, as compared to when only organic results are presented?
- What would time-series analyses of search engine results, that is, monitoring how different search engines

change their approaches towards different types of relevance, show?

- What relevance conflicts can be identified as constitutive of the informational texture of issues as they present themselves through search engines?
- What methods can be developed, considering that much of the relevant data is beyond our control?
- How can alternative approaches to indexing web content and making it searchable be established (e.g., through an Open Web Index, cf. Lewandowski, 2019)?

This list is certainly not exhaustive. We hope that others will add to it. We are looking forward to working together with the research community in taking on these and other crucial questions concerning the societal relevance of search engines.

ENDNOTE

[1] Thank you to Alison Gerber who came up with this apt depiction in a Twitter conversation.

REFERENCES

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

Golebiewski, M., & Boyd, D. (2018). *Data voids: Where missing data can easily be exploited.* Retrieved from https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Data_Voids_Final_3.pdf

Google (2021). *Search engine optimization (SEO) starter guide.* Retrieved from https://developers.google.com/search/docs/beginner/seo-starter-guide?hl=en

Google (n.d.). *How search algorithms work.* Retrieved from https://www.google.com/intl/en_uk/search/howsearchworks/algorithms/

Haider, J., & Sundin, O. (2019). *Invisible search and online search engines: The ubiquity of search in everyday life.* Routledge.

Hillis, K., Petit, M., & Jarrett, K. (2012). *Google and the culture of search.* Routledge.

Hjørland, B. (2010). The foundation of the concept of relevance. *Journal of the American Society for Information Science and Technology*, 61(2), 217–237. https://doi.org/10.1002/asi.21261

Krafft, T. D., Gamer, M., & Zweig, K. A. (2019). What did you see? A study to measure personalization in Google's search engine. *EPJ Data Science*, 8(1), 38. https://doi.org/10.1140/epjds/s13688-019-0217-5

Lewandowski, D. (2012). Credibility in web search engines. In M. Folk & S. Apostel (Eds.), *Online credibility and digital ethos* (pp. 131–146). IGI Global. https://doi.org/10.4018/978-1-4666-2663-8.ch008

Lewandowski, D. (2017). Is Google responsible for providing fair and unbiased results? In M. Taddeo & L. Floridi (Eds.), *The responsibilities of online service providers* (pp. 61–77). Springer. https://doi.org/10.1007/978-3-319-47852-4_4

Lewandowski, D. (2019). The web is missing an essential part of infrastructure: An open web index. *Communications of the ACM*, 62(4), 24–27. https://doi.org/10.1145/3312479

Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, 15(5), 769–787. https://doi.org/10.1080/1369118X.2012.676056

Mager, A. (2018). Internet governance as joint effort: (Re) ordering search engines at the intersection of global and local cultures. *New Media & Society*, 20(10), 3657–3677. https://doi.org/10.1177/1461444818757204

Merriam-Webster. (2021). *Relevance.* Retrieved from https://www.merriam-webster.com/dictionary/relevance

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, 90, 298–307. https://doi.org/10.1016/j.chb.2018.07.043

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* NYU Press.

Nolin, J. (2009). "Relevance" as a boundary concept: Reconsidering early information retrieval. *Journal of Documentation*, 65(5), 745–767. https://doi.org/10.1108/00220410910983092

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you.* Viking.

Rieder, B., & Sire, G. (2014). Conflicts of interest and incentives to bias: A microeconomic critique of Google's tangled position on the Web. *New Media and Society*, 16(2), 195–211. https://doi.org/10.1177/1461444813481195

Roberts, S. T. (2019). *Behind the screen.* Yale University Press.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6), 321–343. https://doi.org/10.1002/asi.4630260604

Saracevic, T. (2016). The notion of relevance in information science: Everybody knows what relevance is. But, what is it? In *Synthesis lectures on information concepts, retrieval, and services.* Morgan & Claypool.

Schultheiß, S., & Lewandowski, D. (2021). Misplaced trust? The relationship between trust, ability to identify commercially influenced results, and search engine preference. *Journal of Information Science.* https://doi.org/10.1177/01655515211014157

SEO Powersuite. (2020). 8 major Google algorithm updates, explained. *Search Engine Land* [Sponsored content]. Retrieved from https://searchengineland.com/8-major-google-algorithm-updates-explained-282627