

Chapter 11

Google Scholar's Filter Bubble: An Inflated Actuality?

Ke Yu

Human Sciences Research Council, South Africa

Nazeem Mustapha

Human Sciences Research Council, South Africa

Nadeem Oozeer

SKA South Africa/AIMS/NWU, South Africa

ABSTRACT

This chapter investigates the allegation that popular online search engine Google applies algorithms to personalise search results therefore yielding different results for the exact same search terms. It specifically examines whether the same alleged filter bubble applies to Google's academic product: Google Scholar. It reports the results from an exploratory experiment of nine keywords carried out for this purpose, varying variables such as disciplines (Natural Science, Social Science and Humanities), geographic locations (north/south), and levels (senior/junior researchers). It also reports a short survey on academic search behaviour. The finding suggests that while Google Scholar, together with Google, has emerged as THE dominant search engine among the participants of this study, the alleged filter bubble is only mildly observable. The Jaccard similarity of search results for all nine keywords is strikingly high, with only one keyword that exhibits a localized bubble at 95% level. This chapter therefore concludes that the filter bubble phenomenon does not warrant concern.

INTRODUCTION

This chapter originates from the publication of the 2011 TED talk by Eli Pariser (*Beware online 'filter bubble'*) and some subsequent exchanges among researchers of anecdotal observations that searches on Google do return varied search results. In an attempt to understand this issue, we started investigating the features and determinants of Google's personalising algorithms. As researchers from the Southern Hemisphere (from now on referred to as South), we asked whether such filter bubbles might impact

DOI: 10.4018/978-1-5225-0830-4.ch011

on our own research processes. We wanted to understand whether the most important personalising determinant—location—meant that our access to knowledge produced elsewhere would be restricted. Moreover, we wanted to understand whether this would influence the quality of our research due to potential variance in our search results. Ultimately, we also wanted to understand whether it would impact the potential success to make our voices heard and compete with scholars from the Northern hemisphere (North) on an equal footing in the global knowledge production.

We decided to investigate these questions by conducting experiments on the extent of the alleged bubbles as well as a survey to understand researchers' academic search behaviour regarding *Google* products. We only focussed on the use of *Google Scholar* for this experiment, as a generic Google search often produces advertisement related information instead of academic information that we are interested. In this chapter, we report on the results from this exploratory experiment, varying variables such as disciplines (Natural Science, Social Science and Humanities), geographic locations (north/south), and levels (senior/junior researchers). We also report on the results from the short survey on academic search behaviour that accompanied the experiment².

The chapter is organised as follows. The introduction outlines the arrival of the information age and the emergence of Google within this context. We also provide some background on Google and Google Scholar's usage in the academic environment and the findings on comparing search effectiveness from Google and other academic search engines. We then turn to the main focus of this chapter on Google Scholar's alleged filter bubbles. We provide a brief review on other studies investigating the extent of Google's bubbles, the methodology we used for our experiments, our design and method of analysis. Then we report our results, as well as that from the search behaviour survey. We conclude by discussing the limitations of the study and implications of our findings.

LITERATURE REVIEW

The Arrival of an Information Age and Its Impact on Information Searching

Over the last two decades, the traditional mode of data, information and knowledge production, as well as storage, have undergone a dramatic revolution. They are no longer in the hands of the few and privileged, but have spread to the greater masses. This process, aided by the emergence and spread of the Internet and web 2.0 tools (for example, wikis, blogs, podcasts and so on) has unleashed a similar revolution in the research world called Science 2.0 or Research 2.0, referring primarily to using the platform of the internet to explore new forms of networking, collaboration, free sharing and open access. Research 2.0 calls on researchers and enables them “to create, annotate, review, re-use and represent information in new ways and of promoting innovations in scholarly communication practices—e.g. publishing work in progress and openly sharing research resources.” (Procter et al., 2010, p. 4039)

The amount of data, information and knowledge produced has exploded in this process. A study by IBM in 2011 reported that the digital data created every 2 days is equivalent to the total amount of data created before 2003. Within academia, Plume and van Weijen (2014) note that the number of articles published has also grown consistently from 1.3 million in 2003 to 2.4 million in 2013³. The explosion of the amount of data/information/knowledge, their virtually limitless availability and accessibility, as well as the ‘retreat’ of the traditional gatekeeper (both for content production and dissemination), is no doubt a blessing. However, it is also a double-edged sword. The increased size and complexity of the

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/google-scholars-filter-bubble/167444

Related Content

Melbourne's Advanced Rail Transportation: Innovative Systems and Their Future Perspective

Koorosh Gharehbaghi, Ken Farnes and Matt Myers (2020). *International Journal of Strategic Engineering* (pp. 24-36).

www.irma-international.org/article/melbournes-advanced-rail-transportation/255140

Delphi MIX for the Construction of Future Critical Scenarios in Social Distancing

Antonio Tintori and Giulia Ciancimino (2022). *Handbook of Research on Advanced Research Methodologies for a Digital Society* (pp. 390-403).

www.irma-international.org/chapter/delphi-mix-for-the-construction-of-future-critical-scenarios-in-social-distancing/287473

Hole Drilling Route Optimization in Printed Circuit Boards Using Far-to-Near Metaheuristics:

Optimizing the Hole Drilling Route via Far-to-Near Metaheuristic

Souhail Dhoub (2022). *International Journal of Strategic Engineering* (pp. 1-12).

www.irma-international.org/article/hole-drilling-route-optimization-in-printed-circuit-boards-using-far-to-near-metaheuristics/301568

Hole Drilling Route Optimization in Printed Circuit Boards Using Far-to-Near Metaheuristics:

Optimizing the Hole Drilling Route via Far-to-Near Metaheuristic

Souhail Dhoub (2022). *International Journal of Strategic Engineering* (pp. 1-12).

www.irma-international.org/article/hole-drilling-route-optimization-in-printed-circuit-boards-using-far-to-near-metaheuristics/301568

Considering the Source: Sampling and Data Collection in a Mixed Methods Study

Mindy Crain-Dorough (2020). *Applied Social Science Approaches to Mixed Methods Research* (pp. 54-85).

www.irma-international.org/chapter/considering-the-source/244115