

Exploring Strategies to Prevent Harm from Web Search

Steven Zimmerman
szimme@essex.ac.uk

Abstract

Web search, the process of seeking and finding information online, is an ubiquitous activity engrained in the lives of many individuals and much of broader society. This activity, which has brought many benefits to individuals and society, has also opened the door to many harms, such as echo chambers, loss of privacy and exposure to misinformation. Members of the information retrieval (IR) community now recognize the dangers of the search technologies commonplace in our daily lives. The upshot of this recognition are growing efforts to address these dangers by the IR community. These efforts focus heavily on system oriented solutions, but give limited focus on behavioural and cognitive biases and behaviours of the search and even less attention to interventions designed to address these biases and behaviours. As such, a theoretical framework is proposed, with behavioural and cognitive strategies as a core component of interactive Web search environments designed to minimize harm.

Using the framework as the foundation, this thesis presents a number of offline and online studies to evaluate *nudging*, a popular intervention strategy rooted in the field of behavioural economics, and *boosting*, a successful intervention strategy from the cognitive sciences, as strategies to reduce risk of harm in Web search. The key takeaway from these studies being that both *boosting* and *nudging* should be considered as viable approaches for harm prevention in Web search environments, in addition to pure system and algorithmic solutions. Additional contributions of this thesis include methods of study design for the comparison of multiple paradigms that promote improved decision making, along with a set of evaluation metrics to measure the success of the IR system and user performance as they relate to the harms being prevented. Future research is needed to confirm the effectiveness of these strategies for other types of harms.

Awarded by: University of Essex, Colchester, England **on** February 24th 2021.

Supervised by: Udo Kruschwitz, Jon Chamberlain and Chris Fox.

Available at: <http://repository.essex.ac.uk/29762/>.

Selected Publications

- Steven Zimmerman, Chris Fox, and Udo Kruschwitz. Improving hate speech detection with deep learning ensembles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, LREC 2018, 2018. URL <https://www.aclweb.org/anthology/L18-1404.pdf>.
- Steven Zimmerman, Alistair Thorpe, Chris Fox, and Udo Kruschwitz. Investigating the interplay between searchers' privacy concerns and their search behavior. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, pages 953–956. ACM, 2019a. URL <https://dl.acm.org/doi/10.1145/3331184.3331280>.
- Steven Zimmerman, Alistair Thorpe, Chris Fox, and Udo Kruschwitz. Privacy nudging in search: Investigating potential impacts. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, CHIIR '19, pages 283–287, 2019b. URL <https://dl.acm.org/doi/10.1145/3295750.3298952>.
- Steven Zimmerman, Stefan Herzog, David Elsweiler, Jon Chamberlain, and Udo Kruschwitz. Towards a framework for harm prevention in web search. In *Bridging the Gap between Information Science, Information Retrieval and Data Science - BIRDS*, 2020a. URL https://birds-ws.github.io/birds2020/assets/papers/BIRDS2020_zImmerman.pdf.
- Steven Zimmerman, Alistair Thorpe, Jon Chamberlain, and Udo Kruschwitz. Towards search strategies for better privacy and information. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, pages 124–134, 2020b. URL <https://dl.acm.org/doi/10.1145/3343413.3377958>.