Does it get better with time? Web search consistency and relevance in the visual representation of the Holocaust

Keywords: web search, Holocaust, consistency, relevance, algorithm audit

Extended Abstract

Web search engines play a major role in curating information about the present, but also about historical phenomena. The growing body of research (e.g. Zavadski & Toepfl, 2019; Makhortykh et al., 2022) demonstrates that search engines play an important role in representing and interpreting the past by prioritising specific information sources and content items. However, while doing so, they can be subjected to malperformance which can take both systematic (e.g. unequal visibility of specific aspects of the historical events) or non-systematic (e.g. occasional retrieval of irrelevant content) forms. Such malperformance is particularly concerning in the case of information about mass atrocities (e.g. the Holocaust), where it can interfere with the ethical obligations of protecting victims' memory.

The important constituent of the performance of search engines in terms of representing the past is how such representation changes over time. Search engines revisit their ranking of sources and individual content items to reflect potential changes in their relevance (e.g. Metlzer, 2009); however, there is currently limited understanding of how changes in relevance apply to historical events, in particular the ones for which there are well-established representation practices as in the case of the Holocaust. Achieving such an understanding is essential for answering the following research questions: To what degree changes in what search algorithms treat as relevant can undermine the consistency of representing the Holocaust? What aspects of the Holocaust become more/less visible over time and does their visibility vary depending on the search engine/query language? Are systematic and non-systematic forms of search malperformance influenced by changes in relevance and, if yes, then how? And what is the desired performance of search engines when dealing with Holocaust-related information?

To address these questions, we audited image search results for the query "Holocaust" in Latin and Cyrillic scripts from six Western and non-Western search engines: Baidu, Bing, DuckDuckGo, Google, Yandex, and Yahoo. We use queries in two scripts to investigate how consistent is search performance across languages and whether dissimilarities in the way the Holocaust is remembered in Western Europe and post-Soviet spaces (particularly Russia) are reflected in the outputs' composition. The same reasoning motivated our selection of search engines: we wanted to compare the two most commonly used search engines in Europe (i.e. Google and Bing) with the engines prevalent in Russian and Chinese markets (i.e. Yandex and Baidu) as well as more niche Western search engines (i.e. DuckDuckGo and Yahoo).

To implement the image search audits, we used a virtual agent-based auditing approach which simulates human browsing behaviour (e.g. entering a search query; Ulloa et al., 2022) using the software. We built a network of CentOS virtual machines (100 machines in 2020 and 60 machines in 2021) in the Frankfurt region of the Amazon Elastic Compute Cloud. On

each machine, we deployed two virtual agents: one in the Chrome browser and one in Mozilla Firefox. Each agent consisted of two browser extensions: a tracker and a bot. The tracker collected the HTML of pages visited in the browser and sent them to a storage server. The bot emulated a sequence of actions, such as visiting an image search engine page, entering the "Holocaust" query, scrolling down the result page to load at least 50 images, and cleaning data accessible by the browser and the search engine's JavaScript to prevent earlier searches affecting the subsequent ones.

To analyse data (i.e. all unique image search results retrieved by the agents), we combined historical and qualitative content analysis. Using authoritative information sources (e.g. the U.S. Holocaust Memorial Museum collections), we attributed each image to detect whether it shows a historical episode associated with the Holocaust and, if yes, then when and where the image was made. This information was then used to identify the visibility of specific Holocaust themes (e.g. deportations, life in camps, or post-war commemoration) and Holocaust sites. In the case of non-historical images, our analysis was focused on whether the image makes antisemitic claims or denies the Holocaust as well as whether it makes fun of the Holocaust (e.g. by referring to it in the format of an Internet meme). The analysis was conducted by the two coders who reached high levels of intercoder reliability (i.e. Krippendorff's alpha of 0.84+ for individual coding categories).

Currently, we completed the analysis for the 2020 data collection. Our findings indicate substantial variation in what search algorithms treat as the most relevant visual representations of the Holocaust between individual engines and between queries in the Latin and Cyrillic scripts. Figure 1 demonstrates that for the query in the Latin script, Western search engines tend to prioritise images of the liberated Holocaust camps, while non-Western ones put major emphasis on images of Holocaust memorials and museums. However, for the query in the Cyrillic script, the visibility of individual themes changes profoundly with Bing, for instance, putting the most emphasis on graphic images of murdered Holocaust victims. Similarly, we observe differences in terms of search malperformance: for the query in the Cyrillic script, some Western engines (e.g. DuckDuckGo and Yahoo) are more likely to prioritise images making antisemitic claims and making fun of the Holocaust, whereas for a major Chinese search engine, Baidu, the shift to the Cyrillic script results in the increased amount of outputs which are not related to the Holocaust. In the next step of the analysis, we will expand it to 2021 data and then systematically compare the results.

References

Makhortykh, M., Urman, A., & Ulloa, R. (2022). Memory, counter-memory and denialism: How search engines circulate information about the Holodomor-related memory wars. *Memory studies*, *15*(6), 1330-1345.

Metzler, D., Jones, R., Peng, F., & Zhang, R. (2009). Improving search relevance for implicitly temporal queries. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 700-701).

Ulloa, R., Makhortykh, M., & Urman, A. (2022). Scaling up search engine audits: Practical insights for algorithm auditing. *Journal of information science*, 01655515221093029.

Zavadski, A., & Toepfl, F. (2019). Querying the Internet as a mnemonic practice: how search engines mediate four types of past events in Russia. *Media, Culture & Society*, *41*(1), 21-37.

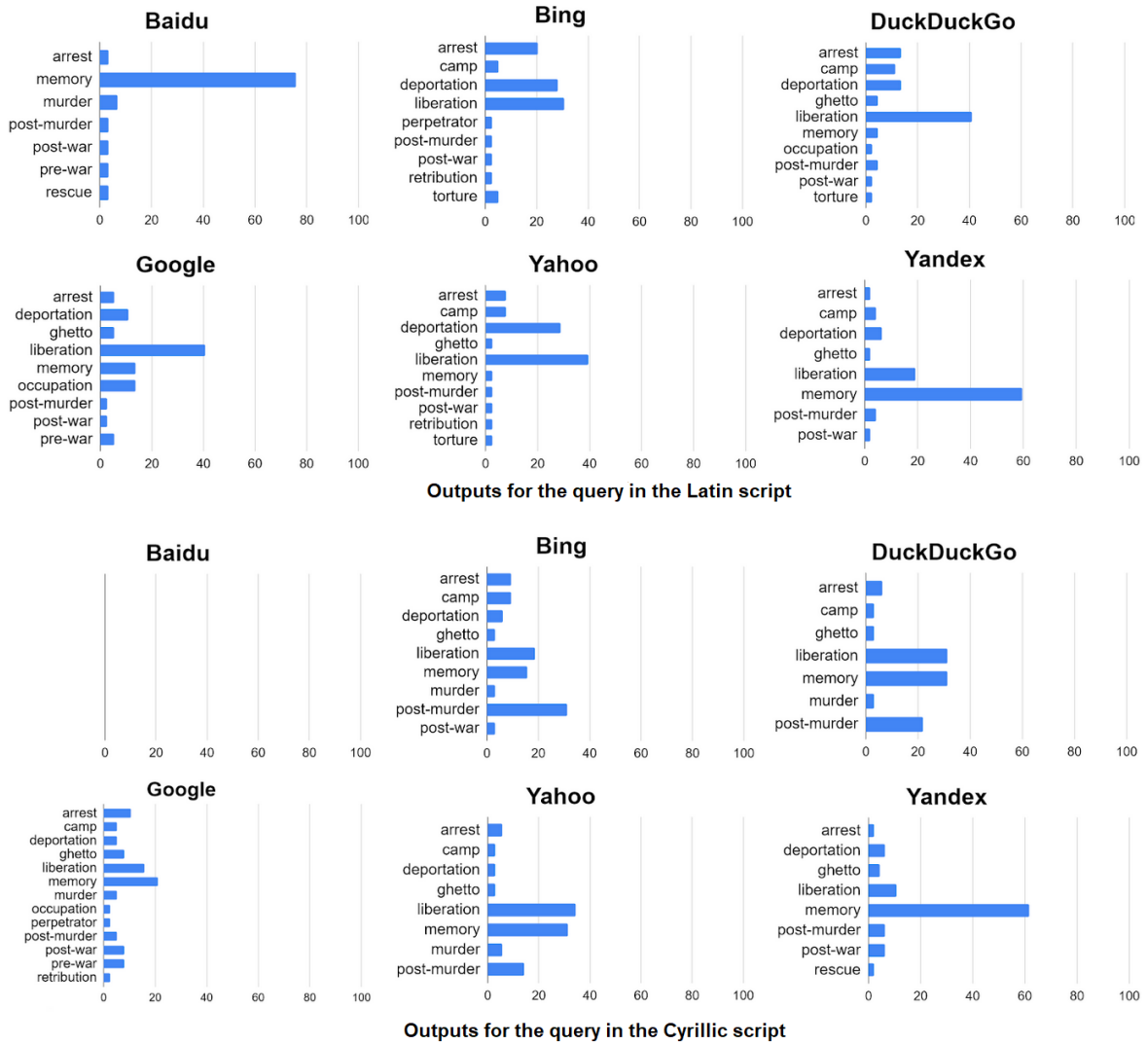

Figure 1. The distribution of image search results for the "Holocaust" query in the Latin and Cyrillic scripts for the six search engines from February 2020 (by the Holocaust theme)