



Extended Abstract

Algorithmic ideology in action: How Google's ranking algorithm impacts different types of information

René König^{1*}

¹ Karlsruhe Institute of Technology, Institute for Technology Assessment and Systems Analysis /
Karlstr. 11, 76133 Karlsruhe, Germany

E-Mail: rene.koenig@kit.edu

* Author to whom correspondence should be addressed; Tel.: +49-721-608-22209; Fax: -

Introduction

Search engines function as important gatekeepers to online information of any kind in more or less every societal domain. Their ranking algorithms determine the visibility of actors and content by creating a hierarchical order of linked websites on the search engine result page (SERP). These algorithmic decisions get additional significance due to the highly concentrated search engine market and the predominant user behavior of only considering the first 10 links on a SERP or even less [2; 7]. In countries like Germany with a stable market share of around 90 % for Google¹, this creates an enormous pressure for websites to be represented within Google's first results for certain keywords. The emerging field of search engine optimization can be regarded as a reaction to this (aiming at achieving a high SERP ranking for websites by adapting them technically to meet the algorithmic criteria for high relevance). Numerous academic publications have also addressed the societal significance of search engines and their wide-ranging information political implications [3; 5; 8-10]. In particular the neutrality of ranking algorithms alleged by Google has been questioned. For example, it has been observed that there are few actors who particularly benefit from the algorithm's favoring of well-linked websites, resulting in a "Googlarchy" [4]. Scholars focusing on the social construction of technology have pointed out search engine developers incorporate specific values in their products leading to an "algorithmic ideology" which serves especially capitalistic needs [6]. However, due to

¹ See: <http://de.statista.com/statistik/daten/studie/167841/umfrage/marktanteile-ausgewaehlter-suchmaschinen-in-deutschland/>.

the secret nature of these algorithms, there is still little knowledge on how exactly they impact online information. To shed light on this, Google rankings for selected queries and algorithm changes have been studied over a period of roughly 5 years.

Methods

In order to understand Google's algorithmic decision-making, suitable queries had to be identified first. For the case studies at hand, the search terms "9/11" (case A) and "climate change" (case B) have been selected. Both promise telling results as they are politically-loaded without implying clear judgment of any kind. Therefore, the results from these queries give insights into the interpretative decisions made by the ranking algorithm (on the contrary, queries like "9/11 conspiracy" or "climate change lies" would lead to rather predictable results which are not very telling in this regard). The Digital Methods Initiative at the University of Amsterdam has automatically queried the selected terms every day in a period of roughly 5 years, collecting the first 100 Google results for each query (A: 06/2007-09/2013, B: 2008-09/2013). Due to technical difficulties the data includes some gaps in which the queries could not be performed. Moreover, the intended method of analysis required a radical reduction of the very large data set: In order to understand what types of websites appeared in the Google results, a qualitative content analysis for each linked website was planned. Therefore, we only selected four days per year (March, June, September, December) and only considered the first 10 results for the content analysis. Since the type of website might have changed over time, for every website the closest version to each selected date was retrieved from the Internet's Archive Wayback Machine (<https://archive.org/web>). This way it was possible to categorize each linked website according to an emerging coding scheme. For example, the website 911truth.org would be coded *CON* for "conspiracy theory" due to its alternative account of the September 11 attacks which differs fundamentally from the website of the 9/11 commission representing the mainstream account of the event (therefore coded *MST*). With this approach, the historical development of content in Google's top ten for the queries could be observed. In a next step, we studied the known changes in Google's algorithm to gain insights into the impact of algorithm changes (see e.g. <http://www.seomoz.org/google-algorithm-change>) in regard to the content represented on the SERP.

Results and Discussion

The results from case A ("9/11") may appear surprising at first sight: The most prominent category was "conspiracy", meaning 34.4 % of all coded websites represented an alternative account of the September 11 attacks (e.g. stating "9/11 was an inside job" by the US government or that the twin towers were brought down by explosives). At the same time, only 15.2 % of the websites were identified as representing the "official" account of the event as it is portrayed in government reports and also by most mass media outlets. A deeper look into the functionality of Google's ranking algorithm makes this result appear less surprising. One of Google's most important ranking factors, the PageRank, regards well-linked websites as more relevant than those sites which received fewer links [1]. Alternative accounts of the September 11 attacks have been actively distributed online by a community called the "9/11 Truth Movement", including websites specifically dedicated to this purpose. We can assume that this community contributed to relatively high PageRanks of such websites by referring to each other via hyperlinks. Additionally, the queried term "9/11" is usually

featured frequently and prominently on these specialized websites. This also helps to be regarded as relevant by Google's algorithm, resulting in a higher ranking.

However, a closer look on the *historical* development of the type of websites in the search engine results gives a more differentiated perspective: While the category CON dominated the SERP for the first years in the given time frame, this drastically changed at the end of 2011: After this point, we rarely found such sites in Google's top ten, whereas the opposing category MST suddenly dominated the results. This became understandable, when we studied Google's algorithm changes. The so-called "Panda update" was introduced exactly at the same time when we observed this drastic switch. It introduced a fundamentally different concept of assessing a website's relevance: Instead of emphasizing the meaning of hyperlinks, now factors like societal acceptance and authority started to play a major role. For example, one of Google's guiding questions to help webmasters achieving a high rank was: "Would you recognize this site as an authoritative source when mentioned by name?"²

Although this correlation cannot with certainty be interpreted as a causation, it appears likely that what was observed was the impact of an algorithm change: While websites representing alternative accounts of 9/11 initially benefited from the emphasis on links, authority became a crucial factor when the Panda update was rolled-out, leading to a higher rank for more conservative sources such as government sites. The presentation will describe these results in greater detail and will also report from case study B which is currently conducted.

Conclusions

The observed patterns reveal how significantly Google's ranking algorithm shapes the type of content that can effectively be accessed through the search engine. It challenges the often expressed expectation of search engines as neutral mediators between the user and the content of the web. Instead we observed that the developers' decisions may lead to a completely different user experience – from one day to the other. Considering Google's important gatekeeping function, it is safe to say that these decisions also have a considerable impact on knowledge societies. Of course, it is still up to the user to transform googled information into knowledge, which is why one should not jump to techno-deterministic conclusions at this stage. However, Google does determine which information can be transformed into knowledge in the first place, as it selects which part of the web we get to see. The historic empirical approach outlined in this paper is an attempt to provide a better understanding of how developers' decisions inscribed in an algorithm concretely impact the user's perception of the web. On a political level, this provokes questions on the lacking transparency of algorithmic decisions: Should users be notified about algorithm updates? How much information on its functionality can a search engine reveal without risking manipulation through search engine optimization? Should governments force search engine providers to create more transparency on their ranking mechanisms? Should users participate in algorithmic decision-making?

² See: <http://googlewebmastercentral.blogspot.nl/2011/05/more-guidance-on-building-high-quality.html>

Acknowledgments

I would like to thank Erik Borra and the Digital Methods Initiative at the University of Amsterdam for providing the data and a number of visualization. Erik was also a great help on technical and intellectual issues.

References and Notes

1. Brin, S.; Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Seventh International World-Wide Web Conference (WWW 1998)*, 14.-18.4. 1998, Brisbane, Australia <http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>.
2. Fallows, D. Search Engine Users. Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve. *Pew Internet & American Life Project*: Washington, USA, 2005, http://www.pewinternet.org/~media/Files/Reports/2005/PIP_Searchengine_users.pdf.pdf.
3. Halavais, A. *Search Engine Society*, Polity Press: Cambridge, UK, 2008.
4. Hindman, M.; Tsioutsoulis, K.; Johnson, J. A. Googlearchy: How a few heavily-linked sites dominate politics on the web. *Annual Meeting of the Midwest Political Science Association*, 31.03.2003, Chicago, USA.
5. Lehmann, K., Schetsche, M.; Eds. *Die Google-Gesellschaft. Vom digitalen Wandel des Wissens*, Transcript Verlag: Bielefeld, Germany, 2007.
6. Mager, A. Algorithmic Ideology. How capitalist society shapes search engines. *Information, Communication & Society* 2012, 15 (5), 769-787.
7. Pan, B.; Hembrooke, H.; Joachims, T.; Lorigo, L.; Gay, G.; Granka, L. In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication* 2007, 12 (3), 801-823.
8. Rogers, R. *Information Politics on the Web*, MIT Press: Cambridge/London, UK, 2004.
9. Röhle, T. Dissecting the Gatekeepers. Relational Perspectives on the Power of Search Engines. In *Deep Search: The Politics of Search Engines beyond Google*, Becker, K., Stalder, F., Eds.; Studienverlag: Innsbruck, Austria, 2009; pp. 117-132.
10. Vaidhyanathan, S. *The Googlization of Everything. And Why We Should Worry*, University of California Press: Berkeley/Los Angeles, USA, 2011.

© 2015 by the authors; licensee MDPI and ISIS. This abstract is distributed under the terms and conditions of the Creative Commons Attribution license.