

René König and Miriam Rasch (eds), *Society of the Query Reader: Reflections on Web Search*, Amsterdam: Institute of Network Cultures, 2014. ISBN: 978-90-818575-8-1.

Historicizing Google Search: A Discussion of the Challenges Related to Archiving Search Results

—

Jacob Ørmen

Historicizing Google Search: A Discussion of the Challenges Related to Archiving Search Results



Jacob Ørmen

Who would not want to know which results you would have gotten if you had entered the keyword ‘terrorism’ into Google’s search bar just before September 11? And what if you could compare it to a search conducted on exactly 11 September 2001 or two weeks after? Or what about tracing the search rankings of websites associated with the query ‘USA’ through the last ten years?

Subjects like these and related types of questions trouble imaginative researchers (probably historians the most), and often the depressing answer presents itself: yeah, you actually could... if someone had just thought of archiving it then. The main issue with historical sources is that someone had to have the idea to produce them in the first place. Luckily such sources have been produced and made available to us in many formats, from newspaper articles, film, and video recordings to official documents and personal correspondence. In recent years we have been able to supplement these types of sources with digital data from social network sites such as Twitter that extend the variety and depth of communication, which we are then able to analyze retrospectively. Now, I argue, we also need to pay attention to those areas of communication that are not as easily stored as newspaper articles or tweets, but nonetheless are relevant sources to future understandings of how events like September 11 or discourses around a particular topic such as ‘USA’ have unfolded. One such area would be Google Search.

Google Search is a particularly interesting case because of its total dominance over the search engine market. Google Search is a central entry point to the web for the majority of people in large parts of the Western world – approximately 65 percent in the U.S. and probably more than 95 percent in Europe.¹ It is an important gateway for people to find information about various topics, events, disasters, etc. and for that reason it is relevant to investigate the type of information that is being presented to individuals in the form of ranked search results. To document the development of search is also important for the general preservation of culture. As Sanz and Stancik put it, Google

1. Ken Hillis, Michael Petit, and Kylie Jarrett, *Google and the Culture of Search*, New York, NY and Oxon, UK: Routledge, 2013.

Search offers us a 'unique empirical window into the study of culture'.² Furthermore, if we want to understand search engines such as Google as a specific 'meta-genre'³ on the internet, it is important that we attend to how these search engines arrange, or mediate, information on the web in the form of ranked search results.

Search results are however not easy to archive since the exact rankings of websites are re-evaluated by proprietary and inaccessible algorithms (including Google's Page-Rank) for each query and thereby subject to change constantly. They are not documents in the same fashion as newspaper articles, tweets, or Facebook posts published online by someone and that appear more or less in a similar way for anyone accessing them. Knowing how a specific website fared or which websites were associated with a particular keyword at any point in time is impossible to assess if the information has not been archived properly. Since the quality of the sources relies heavily on the precise ways they have been archived, these are important issues for scholars in the digital humanities and related fields. Here, I cannot provide a clear answer to the challenges related to archiving search results (since I do not believe there are clear answers here), but I raise some of the most pertinent challenges and suggest some ways forward. Hopefully, this analysis can shed some new light on how search results can be used as historical sources in future research.

'All Those Moments Will Be Lost in Time, Like Tears in Rain'

The problem of irretrievable information is anything but new – in fact, the majority of communication has always been (and still is) lost for eternity. Just think of all oral communication that is not being recorded. It is, as the android Roy Batty so poetically utters in *Blade Runner*, as if 'all those moments will be lost in time, like tears in rain'. This also entails that we of course cannot archive everything (not even all the material on the web), and therefore we must choose carefully what type of information we want to archive and how we want to store it. I believe that search results can serve as important primary sources in the future, and we therefore should worry about which search results merit archiving and how to archive them. Before we can go as far, we need to understand the intricacies of conducting this kind of archiving.

Search engine results exist as a particular type of document online. Since search engines provide an index of retrievable documents on the web, they are on the one hand general access points to a wealth of information (somewhat like a traditional library index). Yet, search engines, by way of various algorithms, also adjust the specific search results to the person making the search (more like the librarian in human form). In that sense the particular search results are a co-creation of the person searching (by the keyword decisions and earlier search history as well as numerous other factors) and Google (by providing an index to search). The results are, so to speak, both 'found' through the index and 'made' by the interaction – referring to discussions of different

-
2. Esteve Sanz and Juraj Stancik, 'Your Search – "Ontological Security" – Matched 111,000 Documents: An Empirical Substantiation of the Cultural Dimension of Online Search', *New Media & Society* (29 April 2013): 0-19.
 3. Klaus Bruhn Jensen, 'Meta-Media and Meta-Communication: Revisiting the Concept of Genre in the Digital Media Environment', *Mediekultur. Journal of Media and Communication Research* 51 (2011): 8-21.

types of data raised by Jensen and others.⁴ They simply don't exist prior to the particular act of searching. Therefore search results are likely to appear differently not only in time and across space as the index and algorithms change, but also between different individuals ('searchers') making the queries. Results do not simply exist 'out there' waiting to be found, scraped, and analyzed as are offline documents, online articles, or tweets, but have to be created in the act of searching. This ontological peculiarity poses a number of unique methodological challenges for researchers studying search results.

The practical question of how to archive Google Search has only become more complicated in recent years. The official API to search through the entire index of Google Search was discontinued by Google as of November 2010 and replaced by a Custom Search API that offers very limited search options.⁵ It appears to be possible to circumvent the limitations set up by Google through a manual customization of the present API, but it remains unclear whether this is in line with Google's Terms of Service (ToS). Furthermore, the present Custom Search API will be discontinued soon (according to rumors on Stack Overflow).⁶ Last but not least, it looks like the search results produced by the search APIs (both the present one and the discontinued) produce quite different search results from manual searches. A different option would be to construct a web scraper and query Google through this device. However, Google has previously explicitly banned this option in the ToS, and it is generally seen as a 'dirty research method'.⁷ Using APIs or server-side access has its clear advantages, but the risk is always that access to or functionalities of services might change suddenly, thereby severely harming the research project. Therefore it is always a risky solution, and even more so for studies taking place over a longer period of time. This is obviously not a unique phenomenon related to studying Google Search, but a common problem that one encounters when retrieving data from online services in general (scholars of Twitter for instance are well aware of the limits the Twitter API sets on research projects).

So it seems that the only really viable option is to collect the results through manual search requests on google.com (and the affiliated sub-domains). An inelegant yet quite feasible method with which to do this is to use research tools that can make automated screen dumps of search results at regular intervals.⁸ This approach has the benefit of containing all the visual information from the browser window, which might work well in research projects that are interested in the constellation of search results (e.g. the size of each result in the query list, the placement of links, images, videos, and other

-
4. Klaus Bruhn Jensen, 'New Media, Old Methods – Internet Methodologies and the Online/Offline Divide', in Mia Consalvo and Charles Ess (eds) *The Handbook of Internet Studies*, Oxford, UK: Wiley-Blackwell, 2010, p. 43.
 5. See, <https://developers.google.com/custom-search/>.
 6. See, 'Google Web Search API Deprecated – What Now?', Stackoverflow, <http://stackoverflow.com/questions/4082966/google-web-search-api-deprecated-what-now>, or 'What Free Web Search APIs Are Available?', Stackoverflow, <http://stackoverflow.com/questions/6084096/what-free-web-search-apis-are-available>.
 7. As Richard Rogers described how people view this kind of research in the plenary session 'The Network Tradition in Communication Research and Scholarship', at the International Communication Association (ICA) Conference, London, June 17-22, 2013.
 8. One such tool is Screenshoter (only for PC), which can take screen dumps of various websites at specified intervals. The program runs in the background and seems to be quite reliable over longer periods of time (e.g. one week). Thank you to Aske Kammer for making me aware of this tool.

contextual data). This method has its obvious drawbacks if the goal is to conduct statistical analysis, since the information is 'flattened' in one image instead of being nicely ordered in a structured database. Manual recoding is of course possible, but will be quite tedious work even with smaller samples of screen dumps. Therefore, screen dumping functions best in more qualitative studies that integrate the visual elements into the analysis. I will discuss the more general issue of why qualitative methods in search engine studies might be more feasible than quantitative later on.

The second important question to consider is how much material is relevant to archive. That of course depends on the purpose for which the material is collected, and there really is no overall answer here. Instead I will present two different archetypical models of archiving that I believe offer alternative ways of approaching search results as documents:

1. *The longitudinal model*: A basic approach would be to archive search results for specific queries on a regular pre-planned basis (e.g. once per week for several years in a row). The point here would be to document certain keywords that retain salience in the popular mind (in other words: that get queried a lot over a longer period of time). Examples of keywords such as this could be 'terrorism', 'United States', and 'E.U.'. Here the goal would be to look for associations between the keywords and websites, pictures, videos, discourses, or anything else one could find as relevant units of analysis. This approach, first of all, makes it possible to engage with how search results of specific terms appear in various stages (in time)⁹ and in various locations (in space) – what Marres and Weltevrede have called the *liveliness* of issues.¹⁰ This approach also entails that we can attend to the contexts of the particular search action and take into account the contextual details of the individual doing the searching (see below). In this way we could conduct what Laura Granka has called 'studies at the micro-level'¹¹ as a supplement to general search trends at the macro-level. However, by creating an archive of specific queries at regular intervals throughout time, researchers could compare the different search result constellations across queries and across time. This way it becomes easier to historicize Google search.
2. *The short burst model*: Another approach would be to collect more material in shorter, yet more intensive, waves of documentation. This could be a relevant method if one suspects that a certain event could cause great disruption in the search rankings. Here the goal is to zoom in on the minute changes happening in the course of the event and then analyze these changes in light of the event (e.g. by relating it to the various sub-events happening or contrasting it to news media coverage, social media activity, etc.) The (hypothetical) documentation of September 11 would be an

9. A perfect example of this is Eric Borra and René König's longitudinal study of which websites were associated most prominently with the search term '9/11' in Google Search across a six year period. See a summary of their conference presentation here: Catalina Iorga, 'Erik Borra and René König Show Google Search Perspectives on 9/11', Institute of Network Cultures, 11 November 2013, <http://networkcultures.org/wpmu/query/2013/11/11/erik-borra-and-rene-konig-google-search-perspectives-on-911/>.

10. Noortje Marres and Esther Weltevrede, 'Scraping the Social? Issues in Real-Time Social Research', *Journal of Cultural Economy* 6.3 (2012): 313-335.

11. Laura A. Granka, 'The Politics of Search: A Decade Retrospective', *The Information Society* 26.5 (September 27, 2010): 364.374.

example of this. By archiving search results before, during, and after certain influential events have occurred (e.g. major pre-planned spectacles such as elections, or sudden disruptive incidents such as natural disasters), we are able to investigate the fluctuations in search results during big events. In this sense the documentation of search results can be used in natural experiment studies, where the event is treated as the exogenous shock. These types of experiments have been used beforehand in the context of Google Search, but there the focus has been on the relationship between click behavior and online advertisement campaigns for certain keywords.¹² However, since we don't have access to Google's algorithms, it is impossible to isolate variables and thereby very difficult to establish a causal link between exogenous events and search result rankings. Nonetheless, this model can provide an interesting insight into the shifting constellations of search results during important events.

Of course it is possible to combine these models into a hybrid (see for example the Danish web archive¹³), and the models outlined above should be seen more as archetypical approaches to the archiving of search results than precise recipes.

Perhaps the greatest challenge spanning both models is that fluctuations in search rankings are very difficult to explain, since the number of factors informing the search rankings and the individual weight of each factor is impossible for us mortal researchers (read: not employed by Google) to decipher. Dirk Lewandowski has ordered these factors into 'query dependent' (considering the position and order of search terms and relating the search terms to amount and types of relevant keywords in documents) and 'query independent' factors (notably the popularity of web sites, determined among other things by the PageRank algorithm).¹⁴ Here I will add personalized factors, which are all those signals stemming from the individual user such as geographical location, prior search history, behavior on other sites Google is able to monitor, and whether one is logged into services (e.g. Google accounts). Thereby, any changes in the search rankings might be due to changes in the algorithms, updates in the index, as well as individual level and country-specific factors. In the following, I discuss ways to tackle this serious issue and suggest a possible path forward. To guide the discussion I introduce a small case study that was originally intended to provide empirical material for another article, but ended up being the cause for why I decided to write this article instead.

The Cat and the Mouse in the Google Sphere

On 14 October 2012 at about 12:08 MDT, some 38 kilometers above the face of the earth, Felix Baumgartner stepped out of the capsule that had carried him up there and jumped out into the stratosphere, thereby beginning his four minute long free-fall towards the ground. After about 40 seconds he reached a top speed of about 1,342 kilometers per hour (or about the height of the Empire State Building per second)

12. Thomas Blake, Chris Nosko, and Steven Tadelis, 'Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment', The National Bureau of Economic Research Working Paper, 6 March 2013, <http://conference.nber.org/confer/2013/EoDs13/Tadelis.pdf>.

13. Danish web archive (netarkivet.dk) archives certain culturally and politically important websites on a routine basis and then archives an extensive number of websites decided on an *ad hoc* basis for specific pre-planned or suddenly occurring events. In that way they combine the models presented here.

14. Dirk Lewandowski, 'Web Searching, Search Engines and Information Retrieval', *Information Services & Use* 25 (2005): 137-147.

breaking the speed of sound before he descended to the ground and landed safely minutes afterwards. The event (named the 'Red Bull Stratos' after its main sponsor) was followed by millions through simultaneous live streams on the web (YouTube alone reported more than seven million viewers at its peak moments) and on Discovery Channel (which obtained the highest ratings for a non-primetime program ever). On Twitter most of the trending hashtags throughout the event were related to the Stratos, and on Facebook Baumgartner's fan page received about half a million new likes and a plethora of comments from ecstatic fans. Even though the global significance of the event might be questionable (cynics might be tempted to re-phrase Neil Armstrong's famous line into 'A giant leap for a man, one small step for mankind') the Stratos was truly a huge media event – though not necessarily in the terms of Dayan and Katz's now classical definition of the genre¹⁵ – that happened across platforms. More importantly for this context it provided a clear-cut case for the study of Google Search (roughly following the short burst model) and an even better case for discussing issues of archiving the searches in real time. The latter proved to be a frustrating yet enlightening game of cat and mouse, which started with the selection of relevant search queries.

Finding the Right Key

I decided to map the Red Bull Stratos about a week before the event took place because it had already received quite a lot of attention from established news media at that point. So I figured that the Stratos could be an important event to document for future research. Obviously, when one tries to document an event like this through Google Search the exact keywords used as search queries are of the greatest importance, since they determine the exact angle taken on the subject. Therefore the keywords must be chosen with care and consideration.

Before the event occurred I couldn't know for sure which queries would be most relevant in a research context, so I tested a number of different keywords in the days leading up to the jump. I eventually decided to map two different keywords: 'red bull stratos' (the official name of the jump) and 'Felix Baumgartner' (the name of the jumper). I also tried with more general keywords such as 'jump', 'stratos', 'felix', and 'baumgartner', but they proved to be too general by including clearly irrelevant search results for the purpose of mapping the event. Therefore I decided to stick with the more precise keywords, which of course meant that I excluded many searches (obviously many people looking for information about the event would use different search terms). Faced with this issue, I decided that false negatives (excluding relevant results) were the better option in this particular context than false positives (including too many irrelevant results).

15. Media (read: TV) events are defined as the pre-arranged 'high holidays of mass communication' that interrupts the daily routines, monopolizes media communication, and encapsulate viewers across the nation and world. Media have the power to unify and speak the language of social integration and reconciliation. (Daniel Dayan and Elihu Katz, *Media Events: The Live Broadcasting of History*, Cambridge, Mass.: Harvard University Press, 1992). Even though the Red Bull Stratos might showcase some of these characteristics, it is unclear whether the event could be said to dominate the media's attention and encapsulate viewers on the same emotional level as traditional media events (the obvious – and very problematic – comparison would be the moon landing in 1969).

By comparing the search volume on Google Trends for the two queries it is very clear that 'Felix Baumgartner' had a greater resonance as a keyword in the population of searchers.¹⁶ It is also clear that the popularity of the search terms are greatest in the Central European countries, particularly Austria, even though the event took place in air space above Nevada, U.S. Since Felix Baumgartner is Austrian, the fact that Austria is the most popular place is not that surprising, but these numbers suggest that there would be a point in looking at various country-specific Google domains in Europe as well.¹⁷

So it seems that the best solution with search queries is either to zoom in on a specific keyword one suspects will retain its relevance over the course of the study (e.g. a very general keyword) or to map different keywords that together can encapsulate the different aspects of the event. This challenge is very much related to the difficulties of determining the proper hashtags to find debates about certain events on Twitter. When comparing queries for 'Red Bull Stratos' and 'Felix Baumgartner' at the same time slot, it is quite obvious that these two keywords capture very different aspects of the event. 'Red Bull Stratos' associates the event much more with official sources (including Red Bull itself), whereas a search for the astronaut yields more person-focused results (among other things, his Facebook page). Similar to how exact hashtags in Twitter research determine the type of debate you can capture, the keywords will demarcate Google Search studies. The words have to be chosen with care. Here, pilot studies (simply trying different keywords in different contexts) and data from Google Trends are a tremendous help.

Who Is the Subject and Who Is the Researcher?

The hard part about studying Google Search on a systematic basis is that Google soon realizes that it is being studied. Even though the so-called 'Hawthorne effect' (the fact that human subjects are conscious of researchers studying them and adapt their behavior accordingly) has been contested in subsequent studies,¹⁸ it might be relevant to ask in the context of studies of Google (and similar companies): to what extent do these services adapt to us studying them? As I see it, this adaption can either be in a very indirect manner, manifesting in the way the search results are shown, or it can be very direct if the search engine intervenes in the study. An example of the latter is shown in Figure 1, where the search engine detects suspicious behavior from the IP address and blocks further requests from that particular computer for a while.

Ethan Zuckerman has described two important ways Google might indirectly interfere. In the first, when Google receives a number of queries from the same IP address, it might try deliberately to randomize search results a bit in order to mask the workings of the algorithm. Zuckerman states, 'The faster you poll the engine, the more variability

16. Comparison made with the free tool Google Trends. Available at: www.google.com/trends/.

17. I actually did exactly that, but because of the issues with personalization the results from the country specific Google domains were rendered more or less meaningless. The language settings simply overruled the specific domain, and I was redirected from the Austrian version of Google to the Danish or U.S. version.

18. Steven D. Levitt and John A. List, 'Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments', *American Economic Journal: Applied Economics. American Economic Association* 3.1 (January 2011): 224-238, <http://www.nber.org/papers/w15016>.

you get, making it harder to profile the engine's behavior'.¹⁹ The second is experimentation. Google is constantly conducting tests (e.g. A/B comparisons) to detect which kind of search results (and design elements) users are most likely to interact with. Because of these issues it becomes very difficult to establish why something is placed at a particular position in the rankings.

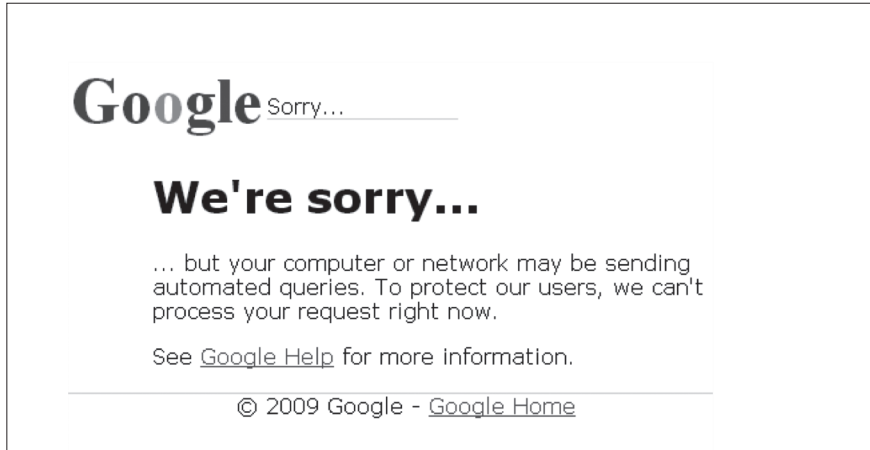


Fig. 1. Google Search error message.

The question we need to ask ourselves as researchers is: How can we study Google when Google is studying us? The only answer I can provide to this question is that we need to be aware of the settings used when searching, then accept that in the end we can never really know the causal relationship involved.

We Know Where You Are... and What Language You Speak

Two particular ranking factors that we need to be acutely aware of are the language settings and IP address. Figure 2 shows the outcome of an attempt to query 'Felix Baumgartner' on google.at (the Austrian version of Google Search) to see the event from an Austrian perspective. Even though I specifically tried to avoid particular Danish search results by doing this, Google still placed these results prominently. There was a video from the Danish-language version of Redbull's website (redbull.dk) as well as a news story from the largest Danish TV channel (nyhederne.tv2.dk). Furthermore, the language settings in the panel on the left side remained Danish. Apart from that, there was only one site in German (wikipedia.de), which seems rather suspicious, especially given the huge attention the event garnered in Austria. Since I did not change my IP address to a server in Germany and had Danish as my default language setting, it was quite likely that these factors informed the search engine's decision to provide me with these search results.

19. Ethan Zuckerman, 'In Soviet Russia, Google Researches You!', ...My Heart's in Accra blog, 24 March 2011, <http://www.ethanzuckerman.com/blog/2011/03/24/in-soviet-russia-google-researches-you/>.

Søgning Billeder Maps Play Gmail Dokumenter Kalender Oversæt Mere -

Google

Søg Ca. 61.400.000 resultater (0,22 sekunder)

Internet **Tip: Søg efter resultater på Dansk alene. Du kan ændre dine sprogindstillinger i Indstillinger**

Billeder

Maps

Videor

Mere

Søg i nærheden af...




Vis søgeværktøjer

Felix Baumgartner
www.felixbaumgartner.com/ - Oversæt denne side
 Ready. Set. Go. DEATH. **FELIX. BAUMGARTNER.** "LEARN TO LOVE WHAT YOU'VE BEEN TAUGHT TO FEAR". -CARL BOENISH. EVERYTIME YOU GEAR UP ...






Felix Baumgartner - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Felix_Baumgartner - Oversæt denne side
Felix Baumgartner (born 20 April 1969) is an Austrian skydiver and a BASE jumper. He set the world record for skydiving 127,000 feet in the sky on October 14, ...
 Biography - See also - References - External links

Felix Baumgartner - Wikipedia
de.wikipedia.org/wiki/Felix_Baumgartner - Oversæt denne side
Felix Baumgartner (* 20. April 1969 in Salzburg) ist ein österreichischer Base-Jumper und Extremsportler. Inhaltsverzeichnis. 1 Leben; 2 Sprung aus der ...

Videor til Felix Baumgartner - Rapporter videoer

		
Felix Baumgartner's Test Jump - Red ... youtube.com 15. mar 2012	Manden bag mission - Felix Baumgartner redbull.dk 6 dage siden	Life on the Edge - Felix ... youtube.com 4. okt 2012

Billeder af Felix Baumgartner - Rapporter billeder

				
		4 dage siden	2 timer siden	8 timer siden

VIDEO: Her springer Felix Baumgartner - TV 2 Nyhederne
nyhederne.tv2.dk/.../id-58634577-video-her-springer-felix-baumgartner...
 20 minutter siden - Se videoen hvor østrigske **Felix Baumgartner** skriver historie og springer ud fra 38,9 km højde.

Fig. 2. Search query 'Felix Baumgartner' on google.at.

To get a clearer idea of whether the language setting of the IP address influenced the constellation of search results, I conducted a mini-experiment (see Figure 3). Here I queried 'Felix Baumgartner' on Google.de (German domain) with four different settings: one with the language set to Danish with my normal IP address in Copenhagen, Denmark (Figure 3a); one with the language set to German with my home IP address in Denmark (Figure 3b); one with the language set to German and with a German IP address (Figure 3c); and one with the language set to Danish and with a German IP address (Figure 3d). The greatest changes in the organic search results seemed to come from the language settings. Notice for example how the country domains on Wikipedia follow the language settings and not the IP address. Meanwhile, the IP address informs the type of ads that are shown to the user in the top banners. So if one wants to appear as if coming from another country when searching Google, it is not enough simply to change the IP address. At a minimum it is required to change the language setting accordingly.



Fig. 3. Search on google.de with varying settings for language and IP-address.

Personalization – the Known Unknowns of Search

Apart from the choice of keywords, language settings, and location-based results, there is of course the increasingly 'black-boxed' issue of personalization. Earlier a number of studies tried to 'second-guess' Google's search algorithm(s) through the systematic mapping of search rankings across queries.²⁰ In recent years it has become increasingly clear that the multitude of factors that informs the exact constellation of search results for any given query²¹ as well as the increasing personalization of users

20. See e.g. Benjamin Edelman, 'Hard-Coding Bias in Google "Algorithmic" Search Results,' 15 November 2010, <http://www.benedelman.org/hardcoding/>.

21. Granka, 'The Politics of Search: A Decade Retrospective'; Zuckerman, 'In Soviet Russia, Google Researches You!'

have made this task very hard, if not impossible.²² There simply exists no vantage point from which the researcher can analyze the search results objectively. We know that the results are personalized (partly because Google confirms this repeatedly),²³ but we don't know exactly how. This fact forces scholars to give up on these strict quantitative designs and either abandon search studies altogether (as I sense quite a few have chosen to do) or work with these limitations actively in their studies.

Richard Rogers, among others, has suggested one way to mitigate the effects of personalization. Rogers advises researchers to operate with a 'research browser', which is a browser cleaned of any history of prior usage and boots directly from scratch each time it is opened.²⁴ Another method commonly applied is to use some kind of IP scrambler that changes the IP address to a random or specified IP address (e.g. through VPN servers) or disguises the IP address (e.g. through TOR). Both of these are definitely viable ways to deal with personalization issues (not to mention surveillance issues) since they make it easier to distort some of the factors that inform the search engine. From a research perspective they also have one downside, as I see it: they run the risk of being too artificial and detached from real-world search situations. Most people are either logged into Google when searching (knowingly or unknowingly), or they do not change their IP address each time to avoid the prior search history to inform their present results. Accordingly, with this approach it might be possible to strip the search engine of some personalized factors and achieve more stable search results across various researchers. The results might be reliable, but not necessarily very valid.

Another way would be to discard strict notions of reliability and embrace search results more as documents intended for qualitative research than precise data points to be used in statistical studies. In this way it might be more appropriate to discuss the scientific value of these types of documents in terms of 'trustworthiness, rigor and quality'²⁵ and to triangulate the queries across searchers (possibly employing human subjects as participants in this process). By these standards would we be able to discuss properly the changes in search results in a more sophisticated manner? Naturally, it would still be virtually impossible to assess which changes in search results are the outcome of personalization and which changes are due to numerous other factors, such as A/B tests and randomization of search results, that are included in the algorithm. But assessing these changes would not be the goal of such a study either. By treating search results as historical documents archived at a specific time and place by researchers with more or less clear biases in their approach (here shown concretely in the personalization mechanisms), they operate on the same level as every other type

22. Martin Feuz, Matthew Fuller, and Felix Stalder, 'Personal Web Searching in the Age of Semantic Capitalism: Diagnosing the Mechanisms of Personalisation', *First Monday* 16.2 (1 February 2011), <http://firstmonday.org/ojs/index.php/fm/article/view/3344/2766>; Eli Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, New York, NY: Penguin Books, 2012.

23. An example: search personalization is a topic on Google Support and has its own fairly detailed subpage: <https://support.google.com/accounts/answer/54041?hl=en>.

24. Richard Rogers, *Digital Methods*, Cambridge, Mass.: MIT Press, 2013.

25. Nahid Golafshani, 'Understanding Reliability and Validity in Qualitative Research', *The Qualitative Report* 8.4 (2003): 597-607.

of source we have access to. Sources can illuminate new aspects of a historical situation or period, but they will never be a complete representation of the subject matter. Sources (like data) can never speak for themselves.

In short, because of all the challenges related to archiving shown here, second-guessing Google is probably not the right way to proceed for research. Instead we should use the search results for a different purpose and in context with other information at hand.

So What's Next...?

As historians would know, relying on single sources is a haphazard affair that could lead to dangerous conclusions (who says that this one source is willing to or capable of telling the truth?). The same goes for search results, whether they are collected as short bursts to document specific events or over a longer period of time for longitudinal studies. Alone they are difficult to verify and thereby almost impossible to analyze in a systematic manner. But in conjunction with other sources (e.g. newspaper articles, activity on social network sites, TV coverage, etc.) they can shed some new light on aspects of societal development and important events that would otherwise remain in the dark. Google is still one of the dominant entryways to the web for many people, as shown in the staggering penetration numbers mentioned earlier, and as such can be an important looking glass into the mentality of the day. For this reason it is important to archive the search results and make them available for future studies.

Throughout the discussion of the various methodological challenges related to online archiving, I have presented some ways forward. To sum this up in a more coherent manner, what I am suggesting here is the following: whether one wants to conduct a longitudinal study or follow a short burst model, it is important to compare (or triangulate) the results from various participants, preferably positioned at different geographical places with appropriate language settings, and either from more or less anonymous networks where IP addresses are not tied to individual machines or from the participants' own computers. If browsers from personal computers are used the criteria for the sampling of human participants are an integral part of the setup. This means, among other things, that the researcher has to consider the personal characteristics of the participants, such as age, gender, place of residence, and search habits, when assessing the search results.²⁶ One such design could employ a 'maximum variation sampling strategy',²⁷ where the researcher attempts to compile a pool of participants with characteristics as different as possible according to specified criteria. If search results vary little across this group of participants, one could establish a stronger case for the consistency of these particular results. As such, this approach to archival research resembles many traditional qualitative research designs.

26. Apart from contextualizing search results with content from other media, there is also the possibility of studying the participants themselves as primary objects of research. This would be a more anthropological, rather than historical, take, but nonetheless important. This method could add interesting insights into our understanding of what people see as important keywords and the reasons they provide for querying specific events.

27. Anton J. Kuzel, 'Sampling in Qualitative Inquiry', in Benjamin F. Crabtree and William L. Miller (eds) *Doing Qualitative Research*, Thousand Oaks, CA: Sage Publications, Inc, 1992, pp. 31-44.

If reliable VPN servers are available, then it is possible to conduct geographical stratification by changing the IP address and language settings instead of relying on human participants. This is indeed a more practical solution (you can do it from one computer and control all the aspects of the archiving yourself), so it is probably more feasible for most individual researchers, but from my experience it seems to be difficult to get assurance that this exercise works in practice. I still believe that a group of human searchers are preferable to this solution, since it offers more detailed and analytically fruitful discussions of how search results differ across profiles.

As a final note, the issue of personalization that I have discussed here in relation to Google seems to be spreading rapidly to new areas, e.g. news websites. Huffington Post is already using an algorithmically informed front-page that adapts to the (perceived) interests of the user.²⁸ This poses an obvious challenge to scholars doing content analysis and makes this conundrum unavoidable for an even greater part of online research. If we want to continue to archive and analyze online content, this is an issue we need to face. The only real solution seems to be to adapt research designs to the limitations of the material and embrace the uncertainties we must accept as researchers. I have suggested one way of doing this here. Maybe others can carry the torch a bit further.

References

- Blake, Thomas, Chris Nosko, and Steven Tadelis. 'Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment', The National Bureau of Economic Research Working Paper, 6 March 2013, <http://conference.nber.org/confer/2013/EoDs13/Tadelis.pdf>.
- Dayan, Daniel and Elihu Katz. *Media Events: The Live Broadcasting of History*, Cambridge, Mass.: Harvard University Press, 1992.
- Edelman, Benjamin. 'Hard-Coding Bias in Google "Algorithmic" Search Results,' 15 November 2010, <http://www.benedelman.org/hardcoding/>
- Feuz, Martin, Matthew Fuller, and Felix Stalder. 'Personal Web Searching in the Age of Semantic Capitalism: Diagnosing the Mechanisms of Personalisation', *First Monday* 16.2 (1 February 2011).
- Fishman, Rob. 'Stories You Might Like: Join Our Beta Program to Test HuffPost Recommendations', *The Huffington Post*, 6 January 2011, http://www.huffingtonpost.com/rob-fishman/stories-you-might-like-jo_b_800427.html.
- Golafshani, Nahid. 'Understanding Reliability and Validity in Qualitative Research', *The Qualitative Report* 8.4 (2003): 597-607.
- Granka, Laura A. 'The Politics of Search: A Decade Retrospective', *The Information Society* 26.5 (September 27, 2010): 364-374.
- Hillis, Ken, Michael Petit, and Kylie Jarrett. *Google and the Culture of Search*, New York, NY and Oxon, UK: Routledge, 2013.
- Iorga, Catalina, 'Erik Borra and René König Show Google Search Perspectives on 9/11', Institute of Network Cultures, 11 November 2013, <http://networkcultures.org/wpmu/query/2013/11/11/erik-borra-and-rene-konig-google-search-perspectives-on-911/>.
- Jensen, Klaus Bruhn. 'Meta-Media and Meta-Communication: Revisiting the Concept of Genre in the Digital Media Environment', *MedieKultur. Journal of Media and Communication Research* 51 (2011): 8-21.

28. See e.g. Rob Fishman, 'Stories You Might Like: Join Our Beta Program to Test HuffPost Recommendations', *The Huffington Post*, 6 January 2011, http://www.huffingtonpost.com/rob-fishman/stories-you-might-like-jo_b_800427.html.

-
- . 'New Media, Old Methods – Internet Methodologies and the Online/Offline Divide', in Mia Consalvo and Charles Ess (eds) *The Handbook of Internet Studies*, Oxford, UK: Wiley-Blackwell, 2010.
- Kuzel, Anton J. 'Sampling in Qualitative Inquiry', in Benjamin F. Crabtree and William L. Miller (eds) *Doing Qualitative Research*, Thousand Oaks, CA: Sage Publications, Inc., 1992, pp. 31-44.
- Levitt, Steven D. and John A. List, 'Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments', *American Economic Journal: Applied Economics. American Economic Association* 3.1 (January 2011): 224-238, <http://www.nber.org/papers/w15016>.
- Lewandowski, Dirk. 'Web Searching, Search Engines and Information Retrieval', *Information Services & Use* 25 (2005): 137-147.
- Marres, Noortje and Esther Weltevrede, 'Scraping the Social? Issues in Real-Time Social Research', *Journal of Cultural Economy* 6.3 (2012): 313-335.
- Pariser, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, New York, NY: Penguin Books, 2012.
- Rogers, Richard. *Digital Methods*, Cambridge, Mass.: MIT Press, 2013.
- Sanz, Esteve and Juraj Stancik. 'Your Search – "Ontological Security" – Matched 111,000 Documents: An Empirical Substantiation of the Cultural Dimension of Online Search', *New Media & Society* (29 April 2013): 0-19.
- Zuckerman, Ethan. 'In Soviet Russia, Google Researches You!', ...My Heart's in Accra blog, 24 March 2011, <http://www.ethanzuckerman.com/blog/2011/03/24/in-soviet-russia-google-researches-you/>.