

Search Engines and Social Science: A Revolution in the Making

Filippo Trevisan
School of Social and Political Sciences
University of Glasgow
25, Bute Gardens
Glasgow – G12 8RS, United Kingdom
Tel: +44-(0)141-330-4377
Email: Filippo.Trevisan@glasgow.ac.uk
Web: www.filippotrevisan.net

April 2013

This paper was prepared as part of the “Google Forum UK” project, which was funded by the Economic and Social Research Council - Knowledge Transfer Small Grant ES/I030166/1:

“Civic Consumers or Commercial Citizens?:

Social Scientists Working with Google UK to Better Understand Online Search Behaviour.”

The author wishes to thank Professor Sarah Oates for her help and advice in the preparation of this paper.

Summary

The analysis of public search-engine results can provide important insights for social-science research. Academics are starting to use tools such as Google Trends much more frequently in their research, although there is no central ‘search-engine’ field of inquiry. Rather, academics using search to address social issues are spread across many disciplines, which makes search valuable across fields but not critical to any one particular field in social science. This paper presents an overview of some of this work as well as makes the important point that the greater value of search-engine research lies in using search analytics as tools for broader inquiry rather than examining search engines themselves as the central object of research. This paper was written as a reflection and further research on the nature of search in social, political and economic aspects of society that was part of a series of meetings between British academics and Google UK. This project was funded by a Knowledge Exchange Grant from the UK Economic and Social Research Council.

Key points

- Google Trends is an especially promising tool that could enable academics to explore new questions.
- Ever since the internet became commercially available in the mid-1990s, search engines have played a crucial role in orienting online traffic, distributing content and constructing knowledge.
- While in the past scholars had been more interested in talking about the role of search engines *themselves* in shaping society, this approach is being increasingly complemented by work that focuses on how search engines can augment academic research *in general*.
- It is crucial that academics investigate opportunities and challenges in the centrality of search engines in contemporary informational patterns and social interaction.
- Emerging strands of research using search-engine tools/data productively include politics and public opinion studies; economics and business; public health and epidemiology; and response/reaction to natural disasters.
- Google Trends data offers advantages over traditional social-science methods such as public opinion surveys. It provides enhanced opportunities to study crisis situations as well as the general relationship between offline events and online behaviour.
- Applications such as Google Trends could provide unprecedented opportunities for examining the connections between new and old media.
- The global dimension of Google Trends as well as the geographical filters that can be applied to its output can facilitate international research by providing comparable data at virtually no cost, thus substantially expanding the scope for social-science research across country boundaries.
- Scholars should be aware of the potential drawbacks associated with this emerging methodology, including doubts over data representativeness when generalising from search engine users to an entire population; language differences and “country effects” in relation to search; as well as limited flexibility afforded by Google Trends.

Discussion

How can social researchers best harness the potential of search engines to identify and explore emerging socio-political trends? Which elements of search engines should be approached that have so far not received the attention they deserve in the nascent field of digital methods and what types of questions could this answer? Stemming from the belief that the time is ripe for the incorporation of search-engine data into social science research, these questions formed the basis of discussions at the Google Forum UK, which in turn provided the inspiration for this paper. The Google Forum UK was a series of biannual events that brought together scholars from a broad array of social science disciplines, from anthropology to political science to media studies to economics, with Google's London team between 2010 and 2012. It was founded by Professor Sarah Oates (then at the University of Glasgow) and included the participation of dozens of UK scholars. These meetings provided new scope for discussing how academics could experiment with using search engines as innovative inquiry tools. As pioneering literature exploring search-engine user behaviour pointed out, "on-going collaboration is needed between the commercial Web search companies and academic researchers to continue to identify and track trends in Web search" (Spink and Jansen, 2005: 189). Google Forum meetings could therefore be seen as a useful response to those recommendations and, more broadly, as highly relevant to the centrality of search to informational practices in the 21st century (Hillis, Petit and Jarrett, 2013: 2). As such, the Google Forum UK was helping to create synergies between social science scholars and the search industry, marking a clear departure from the traditional interpretation of search engines as research objects rather than academic inquiry tools.

With a view to building on this experience, this paper draws on the key issues discussed at Google Forum meetings to provide a useful starting point for those interested in developing innovative uses of search engines in social science research. After a brief overview of the main trends in search-engine scholarship, a strong case is presented for the integration of search engines and their publicly available accessory applications in research projects concerned with capturing, interpreting and possibly anticipating emerging socio-political trends. Opportunities and challenges involved in this process are highlighted and discussed, together with a review of recent social science work that has relied on data collected through search engine applications such as Google Trends, Google News and Google Zeitgeist. In particular, Google Trends is brought into the spotlight as an especially promising tool that could enable academics to explore new questions, although it should be approached with some caution. In doing so, it is hoped that this paper will inform the development of new social science methodologies while also alerting technology developers to some of the challenges faced by academics. At the same time, this paper identifies areas in which there may be scope for change and improvement.

The Story So Far: Search Engines as Social Science Objects

Ever since the internet and web browsers became commercially available, search engines have played a crucial role in orienting online traffic, distributing content and constructing knowledge (Van Couvering, 2008). Meanwhile, the practice of searching itself underwent a radical transformation, from browsing lists that resembled traditional telephone directories in the early years of the internet to trusting commercial search engines to automatically select the content that best fits user needs in more recent times. This has been especially the case since the advent of Google in 1998, whose "models of a good search engine, a good search result, and good algorithmic logic have become normalised as *the* industry standard" (Hillis, Petit and Jarrett, 2013: 53). According to some internet theorists and methodologists, the influence of this process has been so pervasive that it reached beyond the relationship between individual users and the internet, generating a "Googlization" trend capable of affecting multiple aspects of economic, social and

political life (Vaidhyanathan, 2011; Rogers, 2009). According to Vaidhyanathan (2011), “we are folding the interface and structures of Google into our very perceptions” (p. 7), which in turn begs the question of whether “anything (or anyone) matter[s] if it (or she) does not show up on the first page of Google results” (*ibid.*). Broadly speaking, these views are backed up by studies of user-preferences. Despite a slight decrease in the use of search engines in very recent years, which resulted from the simultaneous growth of social networking platforms such as Facebook and Twitter as avenues for finding information online (Dutton and Blank, 2011: 21-2), search engine queries still represent the primary way for most users to retrieve and access online content in countries such as the United Kingdom (*ibid.*) and the United States (Purcell, Brenner and Rainie 2012: 3). This is especially the case among young people who have no experience of the internet prior to Google (Purcell et al., 2012), although that does not necessarily imply that they approach search results uncritically (Thornton, 2010).

In the wake of these developments, a large body of literature has flourished around the socio-technical aspects of search engines, which scholars have cast primarily as *objects* of research (Rogers, 2010). That is, scholars had been more interested in talking about the role of search engines *themselves* in shaping society rather than exploring ways in which search engines can augment academic research *in general*. To date, research into search engines themselves has focused especially on the mechanisms that regulate internet search – where the logic underpinning Google’s PageRank algorithm has provided a blueprint for others to emulate – as well as their impact on social, cognitive and informational practices. To describe this research, Zimmer (2010) coined the expression “Web search studies,” which he envisaged as having four main strands: 1) search engine bias; 2) the role of search engines as gatekeepers of online information; 3) the ethics of search; and 4) legal as well as policy implications of search. Although in practice these lines of inquiry overlap a great deal, the first two – search engine bias and their information gatekeeping role – received the greatest amount of attention in social-science scholarship. As such, it is useful to reflect briefly on the central arguments put forward in this literature. In particular, understanding how users acquire information through search engines and what they expect of keyword searches can provide valuable context to the incorporation of search-engine data into social science methods. This can help to ensure that researchers ask the right questions and challenge established conventions if necessary.

The implications of search for online traffic, content popularity and, ultimately, power have been at the centre of “Web search studies” for over a decade (Introna and Nissenbaum, 2000). Although the fact that the web “is anything but a level, unvariegated network” (Halavais, 2009: 59) is widely accepted as a “necessary evil” that makes the internet navigable, several scholars have criticised the way in which “search engines both contribute to the selection of more prominent sites, and in turn are more influenced by them” (*ibid.*). Some have gone as far as to question the very purpose of search, stating that directing users towards specific websites eliminates “the randomness and non-directional qualities of browsing that produce serendipitous encounters with information [that] are essential to innovation” (Hill, Petit and Jarrett, 2013: 60). In particular, Google’s practice of adopting the number of in-links and traffic levels registered for a given website as proxies for its relevance have been at the heart of disagreement between the many who claim that search engines ensure the perpetuation of dominant ideas and strengthen already powerful actors on one side (see for example: Diaz, 2008; Finkelstein, 2008; Reilly, 2008; Rogers, 2004) and those who regard this mechanism as merely fulfilling the demands of users by optimising results for their preferences and search history on the other (Goldman, 2008; 2011). The sceptical perspective was usefully conceptualised by Hindman (2009) and Hindman, Tsioutsoulis and Johnson (2003) in their introduction of the idea of “Googlearchy” to indicate the propensity of search engines in their contemporary form to perpetuate the status quo by favouring established information sources (e.g. traditional mass media outlets, major political parties, multinational corporations, academic

institutions, etc.) while at the same time marginalising alternative narratives. It has been argued that experienced and motivated searchers can mitigate the effects of search engine bias (Halavais, 2009: 88). Nevertheless, the vast majority of users remains constantly exposed to it as research has shown that most people look at increasingly fewer search results, often focusing exclusively on the very first output page (Jansen and Spink, 2005: 371; 2006: 257).

That said, users have not shown great awareness of these dynamics, let alone concern for their potential effect on information pluralism. Instead, they have developed high levels of trust and attachment to major search providers – especially Google – to the point that, when faced with irrelevant results, they tend to question their proficiency in searching rather than the effectiveness of search engines (Hillis, Petit and Jarrett, 2013: 19). If anything, recent survey data has shown that while users worry about the privacy implications of search giants such as Google and Yahoo! collecting personal information to support customised searches and advertising, this does not detract from their “faith” in the trustworthiness of search results (Purcell, Brenner and Rainie, 2012: 10 and 17-8). In light of these considerations, it could be argued that the success of search engines has influenced other major Internet Service Providers (ISPs). For example, this appears to be the case of Facebook, which in late 2012 launched a search API (Facebook Graph: <https://en-gb.facebook.com/about/graphsearch>) among a selected group of users with plans to extend a semantic search service throughout its network in the future. Despite the lack of detailed information about this project, Facebook’s determination to follow this path marked an important extension of “Google-like” logic to a social networking space where so far the norm has been for users to simply “stumble upon” information as it appears in their newsfeeds while search functionalities have been focused on retrieving individual profiles.

In light of these considerations, the continued popularity of search engines invites social scientists to consider them not only as objects of research. It is now time to focus on search engines as *inquiry tools* capable of providing a wealth of social science information, including new types of data that may not be possible to collect through traditional methods. In particular, search information is bound to expose the interests, attitudes and opinions of users in new and powerful ways. This is because the very act of searching is underpinned by the need for information on a given topic. As Scheitle (2011) noted, “if people are concerned or interested in a particular issue, they will be more likely to search for resources, news, websites, discussion boards, and other types of information related to that issue” (p. 287). In addition, an effective integration of search data into social science research would constitute a way to mitigate the claims of those who have questioned the ethics of corporations such as Google for supposedly “exploiting” internet users when collecting personal information from them with a view to turning it into profit rather than contributing to social progress (Gauntlett, 2011: 191-2 and 212-3). Thus, it is crucial that academics, while remaining alert to the search dynamics discussed above, invest in work that can capitalise on the centrality of search engines in contemporary informational patterns and social interaction. What kind of data is available that could enable researchers to pursue this path? What type(s) of questions could be explored? And, finally, how does that compare with both traditional as well as “virtual” methods and should there be a broad distinction between the two? The next section focuses on these issues.

The Methodological Turn: Search Engines as Social Science Inquiry Tools

In their seminal study on search trends and search engine user behaviour, Spink and Jansen (2005) advocated for the development of electronic transaction log analysis as a valuable alternative to surveys “to gain a clearer understanding of the interactions among searcher, content and Web search engine” (p. 36). While this practice gathered some popularity among researchers focusing on “first generation” search engines such as AltaVista (Halavais, 2009: 46), as new providers took hold of the search market difficulties in obtaining this type of data severely restricted the scope and

significance of this type of inquiry. Most notably, search companies such as Google and Yahoo! have typically been extremely protective of user activity logs, both for understandable commercial reasons and due to the ethical implications of releasing such data, which could expose users to privacy breaches or even unjustified surveillance by state as well as non-state actors (Morozov, 2011: 164-5; Halavais, 2009: 145-51). Yet, is the analysis of raw search records truly necessary in order to unlock the potential of search engines for social science inquiry, or are there other, more straightforward ways for scholars to gather and analyse search information that respect both the business model of search providers as well as the privacy of individual users? Discussions at Google Forum meetings challenged the presumed indispensability of activity log data, focusing instead on what researchers could do to take a more creative approach to the conspicuous amounts of information that search engines such as Google have already made available through accessory applications. Several opportunities emerged that clarified how placing excessive emphasis on the need for raw search data had in fact created a self-perpetuating impasse that hindered the development of other lines of inquiry in this very promising area.

Although these issues have yet to be tackled organically in methodological literature, in very recent years several authors have started to investigate user preferences through data drawn from free online tools such as Google Trends (Scharkow and Vogelsgang, 2011), Google News (Weaver and Bimber, 2008) and Google Zeitgeist (Jeong and Mahmood, 2011). These tools have been incorporated into research designs either as “methods” in their own right or in combination with more established techniques such as surveys and content analysis. In particular, Google Trends (www.google.com/trends) has become the focus of growing interest among social scientists since Google released a report that examined the possibility of using this tool to generate estimates for key economic indicators more rapidly and arguably more accurately than traditional government and industry statistics (Choi and Varian, 2009). Google Trends does not grant access to raw search logs. Instead, it generates aggregate data that shows fluctuations in the search volume for any keyword(s) from 2004 onwards provided that a minimum number of relevant searches were carried out. Its output consists of data that is normalised by measuring it against the total number of searches registered during the period under scrutiny. Final scores are obtained by assigning a value of 100 to the day or week (depending on the length of the period being examined) that registered the highest number of relevant searches and calculating all other scores in function of their distance from the top result on a scale from zero to 99. In addition to its longitudinal dimension, Google Trends also includes a series of geographical filters that enable users to organise the data by country, state/region, or city. Furthermore, users can download Google Trends data to build additional visualisations or for further statistical elaboration, including comparing results to relevant indicators obtained through traditional social science methods such as survey questionnaires.

In addition to the reports mentioned above, a number of studies have sought to apply Google Trends data to a broad array of social science disciplines. These include:

- Politics and public opinion studies (Scheitle, 2011; Scharkow and Vogelsgang, 2011; Manzano and Ura, 2013);
- Economics and business (Hand and Judge, 2011; Kaeserbauer, Hohenstatt and Reed, 2012; Schmidt and Vosen, 2012; Smith, 2012);
- Public health and epidemiology (Breyer et al., 2011; Carneiro and Mylonakis, 2009; Carr and Dunsiger, 2012; Connolly, Postma and Silber, 2009; Frijters et al., 2013; Glynn et al., 2011; Gunn and Lester, 2012; Metcalf, Price and Powell, 2010; Wallcott et al., 2011; Zhou, Ye and Feng, 2011);
- Understanding the environment and reaction to natural disasters (Chai and Sasaki, 2011; Sherman-Morris, Senkbeil and Carver, 2011; Van der Velde, 2012; Wilde and Pope, 2013)

This type of research is still very much in its infancy. Thus, the full incorporation of Google Trends data into social-science scholarship is undoubtedly going to require further refinement. Nevertheless, this preliminary literature list already provides a sense for the inter-disciplinary nature of this methodological innovation. In this context, it is crucial for researchers to engage in a wide-ranging discussion about the opportunities as well as the challenges involved in turning search engines from objects of research to inquiry tools. While this is not the place for an in-depth examination of the various statistical techniques employed in the studies cited above, it is instead particularly useful to examine the main trans-disciplinary themes that emerged from both discussions at Google Forum meetings and innovative literature published to date.

Google Trends and Pioneering Research: Opportunities and Challenges

Overall, the potential benefits of Google Trends data for social-science scholarship can be grouped into two main categories. These are:

- 1) Advantages over traditional methods such as public opinion surveys; and
- 2) Enhanced opportunities to study the relationship between offline events and online behaviour, especially during crisis situations.

Conversely, scholars should also be aware of the potential drawbacks associated with this emerging methodology, namely doubts over data representativeness when generalising from search engine users to an entire population; language differences and “country effects” in relation to search; as well as the lack of detail in Google Trends. Each of these is discussed briefly in the paragraphs that follow.

Advantages over traditional social science methods

Using Google Trends to gather information about user interests, concerns and behaviour could have several advantages over traditional methods, especially surveys. While some may find Google’s own claim that this type of data could enable researchers to “predict the present” (Choi and Varian, 2009) excessively optimistic, its ability to reveal emerging socio-political trends quickly and reliably has been tested in the area of public opinion and public mood research. In particular, researchers such as Scheitle (2011) and Scharkow and Vogelsgang (2011) obtained encouraging results when comparing Google Trends data with traditional survey results for issue salience in an effort to construct a reliable “issue barometer.” In addition, Manzano and Ura (2013) have highlighted the possibility that aggregate search-engine data may also enable researchers to uncover information that would remain “hidden” or only be possible to capture at a much greater cost in survey investigations. In their recent study of public opinion over the appointment of U.S. Supreme Court Judge Sonia Sotomayor, Manzano and Ura focused on the geographical distribution of relevant keyword searches as shown in Google Trends to reveal a possible link between people’s interest in this issue and their specific ethnic background. This made for a useful example of how this tool could be used not only to explore population-wide trends, but also to focus on issues of particular interest to specific social groups, including minorities. More broadly, these and several other studies among those listed above emphasise the fact that Google Trends data, which can be obtained at virtually no cost, mitigates research bias and the incidence of incomplete or false responses as it relies on actual search logs rather than requiring users to fill in questionnaires. In addition, it is unobtrusive while also respecting the privacy of users because personal information is removed by Google before data are elaborated and released. That said, doubts have nevertheless been raised over the actual representativeness of this data, which invite some additional considerations.

Representativeness of Google Trends data

Google Trends results are drawn from a comprehensive search log database. Yet, the issue remains whether it would be appropriate for scholars to generalise conclusions obtained through this method to make claims about entire populations. Most notably, this is because of potential differences between search engine users and non-users. So far, research pioneers in this area have postponed the debate on this issue, prioritising methodological innovation over data validity questions. Thus, they limited themselves to assuming that, “unless the underlying mechanisms for seeking information about salient issues are fundamentally different for onliners than for offliners, the validity of online search queries for measuring the public agenda is not at risk” (Scharkow and Vogelsang, 2011: 107). Similarly, other caveats may derive from the fact that, although Google Trends tells us what topics users searched when, it cannot tell us the motives that underpinned such keyword searches. Thus, researchers can only make reasonable assumptions as to why users were interested in a given issue or set of issues. As Carneiro and Mylonakis (2009) noted in their work on tracking disease outbreaks using Google Trends, “naturally, all the people searching for influenza-related topics are not ill, but trends emerge when all influenza-related searches are added together” (p. 1557). Although the comparison they carried out between Google Trends data and official medical records seemed to confirm this assumption, the issue of data representativeness remains a central one that ought to be examined exhaustively if search engine accessory applications are to make further inroads into socio-political research that looks beyond the relationship between internet users and technology.

While this is not the place for an in-depth discussion of digital divide studies and internet usage patterns, it is important to point out that the issue of data representativeness is not exclusive to research that focuses on search engine results. Rather, it also applies to work that relies on data drawn from other online platforms, especially social media. In this context, it is useful to note that, despite ongoing growth in the number of social networking sites users, search engine users continue to constitute both a substantially larger group as well as one that better represents the general population. This is apparent when comparing the latest search engine usage estimates with, for example, Twitter penetration rates in a country such as the United States (where Twitter is relatively popular). In 2012, over 90% of American internet users across all age groups under 65 relied on search engines to retrieve information online (Purcell, Brenner and Rainie, 2012: 5-6). In contrast, only 16% of them used Twitter, which in turn was even less popular among those aged 30 and over (Duggan and Brenner, 2012: 4). In light of these trends, search-engine data such as that provided by Google Trends makes for a substantially less skewed source of information on public mood and issue salience than Twitter records, which in recent years have nonetheless been employed to study people’s reaction to and involvement in a broad array of events, from election campaigns (Tumasjan et al., 2010) to the Arab Spring uprisings (Aday et al., 2012). Undoubtedly, a much broader debate will be needed to dispel all doubts over the representativeness of search-engine data. However, these preliminary considerations make for encouraging premises.

Crisis research and the relationship between online traffic and offline events

Following on from the last point above, a further key application of Google Trends data could be in work that seeks to capture and explain the evolution of crises. In this context, the ability of search engine accessory tools to relay information about user interests nearly in real time could be of crucial importance to those studying public reactions to emergency situations such as natural disasters with a view to supporting more effective responses from the authorities (Chai and Sasaki, 2011; Sherman-Morris, Senkbeil and Carver, 2011). More broadly, search-engine aggregate data also could be useful in analysing the relationship between online behaviour and offline events over

longer time periods. At a most basic level, online-interest charts generated with Google Trends can be compared with a timeline of relevant offline events to verify the existence of links between online traffic for a certain topic and related “real world” developments. This approach was adopted to generate important contextual information in recent work looking at online reactions to press coverage of celebrity issues (Metcalf, Price and Powell, 2010) as well as controversial policy plans in a contentious area such as that of welfare provision (Trevisan, 2013).

In addition, data highlighting fluctuations in keyword searches associated with specific topics or ideas could also provide crucial opportunities for verifying the pervasiveness of public information campaign messages as well as for studying the genesis and evolution of dominant mass media frames in the age of the internet. Although further empirical and conceptual work will be needed to better understand the potential of Google Trends in this emerging research area, some pioneering scholars have reached intriguing conclusions when experimenting with ways to compare search engine records for selected topics to, for example, content analysis of news media material (Metcalf, Price and Powell, 2010) as well as public health information campaigns (Glynn et al., 2011). Overall, this could make for a fundamental line of inquiry at a time in which both news production and consumption occur in a highly hybridised media environment (Chadwick, 2011) where traditional notions of “audience” as information recipients and journalists as news providers have become blurred. Applications such as Google Trends could therefore provide unprecedented opportunities for examining the connections between new and old media, as well as links between online and offline messages that so far have been under-researched. To what extent are dominant media frames or, for example, election campaign messages reflected in online search patterns? What is the relationship between key search trends and the news media agenda? While the potential for methodological innovation in this area is extensive, this process will undoubtedly involve a series of challenges and limitations that researchers should consider very carefully, as discussed in the next section.

Country effects and limited granularity

The global dimension of Google Trends as well as the geographical filters that can be applied to its output can facilitate international research by providing comparable data at virtually no cost, thus substantially expanding the scope for online research. This is especially relevant for the study of online politics, which so far has focused almost exclusively on election campaigns despite calls from scholars for a broader comparative outlook in this area. However, when attempting to draw comparative observations from search-engine data, researchers should keep in mind a problem that may require them to integrate their analysis with additional information from other sources or, in some cases, restrict the scope of international work. In particular, investigations relying on Google Trends as the main inquiry tool should pay close attention to what could be described as “country effects,” i.e. the consequences of Google’s relative position in the search market of the country or countries under scrutiny. With regard to this issue, three typical situations can be identified. First, there are countries in which Google enjoys an unrivalled dominance in the search market such as in the United States, the United Kingdom and other Western democratic nations. At the other extreme there are countries where Google is banned from operating (for example in Iran). Third, there are a series of countries in which Google occupies a secondary position in the search market, which is instead controlled by “home-grown” providers. Most notably, this is the case of China and Russia, where the majority of searches are carried out through local engines Baidu and Yandex respectively (Oates, 2011; Halavais, 2009), neither of which makes aggregate search data publicly available. While researchers have effectively no control over these differences, they must be aware of their important implications for data representativeness.

A further issue that may hinder the type of research proposed above is the limited granularity that can be applied to the analysis of keyword search records with Google Trends. This is likely to create particularly challenging research issues in relation to keywords and phrases with multiple meanings or that may simultaneously apply to several different topics. As noted above, Google Trends provides information about the words and phrases searched by users, yet researchers can only advance reasonable hypotheses with regard to the motives that prompted users to search for a given expression as well as the exact type of information for which they were searching. This potentially opens up the analysis of search-engine data to the influence of irrelevant searches, which can be mitigated only partially by applying fixed “category” filters to Google Trends queries (e.g. “law & government,” “news,” “arts & entertainment,” “health,” etc.). Therefore, the impossibility of identifying and eliminating irrelevant search logs from the data from which Google Trends output is elaborated calls for great care in the selection of keywords. Researchers should avoid as much as possible a focus on expressions that could be applied simultaneously to multiple contexts and different topics. In addition, attention should be paid also to the language selected for investigation as searching for the same term(s) in different languages is likely to yield different results while Google itself generally tends to privilege English language sources (Al-Eroud et al., 2011).

Conclusions

Overall, these considerations, discussions at Google Forum meetings as well as the literature cited in this paper have shown that social-science scholarship is on the verge of a substantial methodological shift through the incorporation of publicly available search-engine data. As new potential applications for Google Trends and other accessory search engine tools continue to emerge, it is therefore crucial for social researchers to tackle both opportunities and challenges in an organic debate through empirical as well as conceptual inter-disciplinary work. Undoubtedly, the issues highlighted in this paper cannot represent an exhaustive agenda for the study of search engines as social science inquiry tools. Rather, this sought to provide useful reference points with regard to this nascent methodological strand by drawing together key themes that emerged from pioneering research and reflecting on some of the issues in the academic literature. It is therefore hoped that, while researchers strive to be more creative in their approach to search-engine data for social-science research, technology developers will simultaneously benefit from this debate and be able to address some of the issues highlighted by academics.

References

- Aday, Sean, Farrell, Henry, Lynch, Mark, Sided, John, and Freelon, Deen (2012) *Blogs and Bullets II: New Media and Conflict After the Arab Spring*, Washington, DC: United States Institute of Peace. Available for free download at: <http://www.usip.org/publications/blogs-and-bullets-ii-new-media-and-conflict-after-the-arab-spring> (accessed: 15th Feb. 2013).
- Al-Eroud, Ahmed, Al-Ramahi, Mohammad, Al-Kabi, Mohammad, Alsmadi, Izzat, and Al-Shawakfa, Emad (2011) “Evaluating Google Queries Based on Language Preferences,” *Journal of Information Science*, 37(3): 282-92.
- Breyer, Benjamin, Saunak, Sen, Aaronson, David, Stoller, Marshall, Erickson, Bradley, and Eisenberg, Michael (2011) “Use of Google Insights for Search to Track Seasonal and Geographic Kidney Stone Incidence in the United States,” *Endourology and Stones*, 78(2): 267-71.
- Carneiro, Herman A., and Mylonakis, Eleftherios (2009) “Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks,” *Clinical Infectious Diseases*, 49(15 Nov.): 1557-64.

- Carr, Lucas, and Dunsiger, Shira (2012) "Search Query Data to Monitor Interest in Behavior Change: Application for Public Health," *PLoS ONE*, 7(10): 1-4.
- Chadwick, Andrew (2011) "The Political Information Cycle in a Hybrid News System: The British Prime Minister and the "Bullygate" Affair," *The International Journal of Press/Politics*, 16(1): 3-29.
- Chai, Sengtha, and Sasaki, Nophea (2011) "Using Online Tools to Assess Public Responses to Climate Change Mitigation Policies in Japan," *Future Internet*, 3: 117-29.
- Choi, Hyunyoung, and Varian, Hal (2009) "Predicting the Present With Google Trends," Google Inc.
- Connolly, Mark, Postma, Maarten, and Silber, Sherman (2009) "What's on the Mind of IVF Consumers?," *Reproductive BioMedicine Online* 19(6): 767-9.
- Diaz, Alejandro (2008), "Through the Google Goggles: Sociopolitical Bias in Search Engine Design," in Spink, A., and Zimmer, M., (eds.), *Web Search: Multidisciplinary Perspectives*, Berlin: Springer, pp. 11-34.
- Duggan, Maeve, and Brenner, Joanna (2012) *The Demographics of Social Media Users – 2012*, Washington, DC: Pew Internet and American Life Project.
- Dutton, Bill, and Blank, Grant (2011) *Next Generation Users: The Internet in Britain*, Oxford: Oxford Internet Institute.
- Finkelstein, Seth (2008) "Google, Links, and Popularity versus Authority," in Turow, J., and Tsui, L. (eds.) *The Hyperlinked Society: Questioning Connections in the Digital Age*, Ann Arbor: The University of Michigan Press, pp. 104-24.
- Frijters, Paul, Johnston, David, Lordan, Grace, and Shields, Michael (2013) "Exploring the Relationship between Macroeconomic Conditions and Problem Drinking as Captured by Google Searches in the US," *Social Science and Medicine*, 84: 61-8.
- Gauntlett, David (2011) *Making is Connecting: The Social Meaning of Creativity, from DIY and Knitting to YouTube and Web 2.0*, Cambridge: Polity Press.
- Glynn, Ronan, Kelly, John, Coffey, Norma, Sweeney, Karl, and Kerin, Michael (2011) "The Effect of Breast Cancer Awareness Month on Internet Search Activity – A Comparison with Awareness Campaigns for Lung and Prostate Cancer," *BMC Cancer*, 11: 442.
- Goldman, Eric (2008) "Search Engine Bias and the Demise of Search Engine Utopianism," in Spink, A., and Zimmer, M. (eds.) *Web Search*, Berlin: Springer, pp. 121-33.
- (2011) "Revisiting Search Engines Bias," *Santa Clara University Legal Studies Research Paper* 12-11.
- Gunn, John, and Lester, David (2012) "Using Google Searches on the Internet to Monitor Suicidal Behavior," *Journal of Affective Disorders*.
- Halavais, Alexander (2009) *Search Engine Society*, Cambridge: Polity.
- Hand, Chris, and Judge, Guy (2012) "Searching for the Picture: Forecasting UK Cinema Admissions Using Google Trends Data," *Applied Economics Letters*, 19(11): 1051-55.
- Hilis, Ken, Petit, Michael, and Jarrett, Kylie (2013) *Google and the Culture of Search*, New York: Routledge.
- Hindman, Matthew (2009) *The Myth of Digital Democracy*, Princeton, NJ: Princeton University Press.
- Hindman, Matthew, Tsioutsoulouklis, Kostas, and Johnson, Judy A. (2003) " 'Googlearchy': How A Few Heavily-Linked Sites Dominate Politics on the Web," paper prepared for the annual meeting of the Midwest Political Science Association, Chicago, IL.
- Introna, Lucas D., and Nissenbaum, Helen, (2000), "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society*, 16(3): 169-85.
- Jansen, Bernard, and Spink, Amanda (2005) "An Analysis of Web Searching by European Alltheweb.com Users," *Information Processing and Management*, 41: 361-81
- (2006) "How Are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs," *Information Processing and Management*, 42: 248-63.

- Jeong, Yongick, and Mahmood, Reaz (2011) "Reading the World's Mind: Political, Socioeconomical and Cultural Approaches to Understanding Worldwide Internet Search Queries," *International Communication Gazette*, 73(3): 233-51.
- Kaeserbauer, Manuel, Hohenstatt, Ralf, and Reed, Richard (2012) "Direct Versus Search Engine Traffic: An Innovative Approach to Demand Analysis in the Property Market," *International Journal of Housing Markets and Analysis*, 5(4): 392-413.
- Manzano, Sylvia, and Ura, Joseph (2013) "Desperately Seeking Sonia? Latino Heterogeneity and Geographic Variation in Web Searches for Judge Sonia Sotomayor," *Political Communication*, 30(1): 81-99.
- Metcalf, D., Price, C. and Powell, J. (2010) "Media Coverage and Public Reaction to a Celebrity Cancer Diagnosis," *Journal of Public Health*, 33(1): 80-5.
- Morozov, Evgeny (2011) *The Net Delusion: How Not To Liberate the World*, London: Penguin.
- Oates, Sarah (2011) "Going Native: The Value in Reconceptualizing International Internet Service Providers as Domestic Media Outlets," *Philosophy & Technology* 24 (4): 391-409.
- Purcell, Kristen, Brenner, Joanna, and Rainie, Lee (2012) *Search Engine Use 2012*, Washington, DC: Pew Internet and American Life Project.
- Purcell, Kristen, Rainie, Lee, Heaps, Alan, Buchanan, Judy, Friedrich, Linda, Jacklin, Amanda, Chen, Clara, and Zickuhr, Kathryn (2012) *How Teens Do Research in the Digital World*, Washington, DC: Pew Internet and American Life Project.
- Reilly, Paul, (2008), "Googling Terrorists: Are Northern Irish Terrorists Visible on Internet Search Engines?," in Spink, A., and Zimmer, M., (eds.), *Web Search: Multidisciplinary Perspectives*, Berlin: Springer, pp. 151-75.
- Rogers, Richard (2004) *Information Politics on the Web*, Cambridge, MA: MIT Press.
- (2009) "The Googlization Question: Towards the Inculpable Search Engine?," in Becker, K., and Stalder, F. (eds.) *Deep Search: The Politics of Search Beyond Google*, Innsbruck: Studien Verlag, pp. 173-84.
- (2010) "Internet Research: The Question of Method – A Keynote Address from the Youtube and the 2008 Election Cycle in the United States Conference," *Journal of Information Technology and Politics*, 7(1): 241-60.
- Scharkow, Michael, and Vogelsgang, Jens (2011) "Measuring the Public Agenda Using Search Engine Queries," *International Journal of Public Opinion Research*, 23(1): 104-13.
- Scheitle, Christopher P. (2011) "Google's Insights for Search: A Note Evaluating the Use of Search Engine Data in Social Research," *Social Science Quarterly*, 92(1): 285-95.
- Schmit, Torsten, and Vosen, Simeon (2012) "Using Internet Data to Account for Special Events in Economic Forecasting," *Ruhr Economic Papers* 382.
- Sherman-Morris, Kathleen, Senkbeil, Jason, and Carver, Robert (2011) "Who's Googling What? What Internet Searches Reveal about Hurricane Information Seeking," *Bulletin of the American Meteorological Society*, August 2011: 975-85.
- Spink, Amanda, and Jansen, Bernard (2005) *Web Search: Public Searching on the Web*, Boston: Kluwer Academic Publishers.
- Thornton, Stephen (2010) "From 'Scuba Diving' to 'Jet Skiing'? Information Behavior, Political Science, and the Google Generation," *Journal of Political Science Education*, 6(4): 353-68.
- Trevisan, Filippo (2013) "Disabled People, Digital Campaigns, and Contentious Politics: Upload Successful or Connection Failed?," in Scullion, R., Lilleker, D., Jackson, D., and Gerodimos, R., (eds.), *The Media, Political Participation, and Empowerment*, London: Routledge, pp. 175-91.
- Tumasjan, Andranik, Sprenger, Timm O., Sandner, Philipp G., and Welp, Isabell M. (2010) "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Available for free download:
<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441/1852>

- Vaidhyanathan, Siva (2011) *The Googlization of Everything (And Why We Should Worry)*, Berkeley: University of California Press.
- Van Couvering, Elizabeth (2008) "The History of the Internet Search Engine: Navigational Media and the Traffic Commodity," in Spink, A., and Zimmer, M., (eds.), *Web Search: Multidisciplinary Perspectives*, Berlin: Springer, pp. 177-206.
- van der Velde, Marijn, See, Linda, Fritz, Steffen, Verheijen, Frank, Khabarov, Nikolay, and Obersteiner, Michael (2012) "Generating Crop Calendars with Web Search Data," *Environmental Research Letters*, 7(2012).
- Walcott, Brian, Nahed, Brian, Kahle, Kristopher, Redjal, Navid, and Coumans, Jean-Valery (2011) "Determination of Geographic Variance in Stroke Prevalence Using Internet Search Engine Analytics," *Neurosurgery Focus*, 30.
- Weaver, David, and Bimber, Bruce (2008) "Finding News Stories: A Comparison of Searches Using Lexisnexis and Google News," *Journalism and Mass Communication Quarterly*, 85(3): 515-30.
- Wilde, G., and Pope, K. (2013) "Worldwide Trends in Fishing Interest Indicated by Internet Search Volume," *Fisheries Management and Ecology*, 20:211-22.
- Zhou, Xichuan, and Ye, Jieping (2011) "Tuberculosis Surveillance by Analyzing Google Trends," *IEEE Transactions on Biomedical Engineering*, 58(8): 2247-54.
- Zimmer, Michael (2010) "Web Search Studies: Multidisciplinary Perspectives on Web Search Engines," in Husinger, J., Klastrop, L., and Allen, M. (eds.) *International Handbook of Internet Research*, Dordrecht: Springer, pp. 507-21.