Selected Papers of Internet Research 16:
The 16th Annual Meeting of the
Association of Internet Researchers
Phoenix, AZ, USA / 21-24 October 2015

AN IMAGINARY ALGORITHMIC PUBLIC: HOW MEDIA REPORT ON SEARCH QUERY METRICS

Anna Jobin
University of Lausanne

Abstract

Web search engines provide search query metrics in different formats for different audiences, notably annual search query statistics by country, which are often reported on by traditional media outlets. These statistics present top-ranking query charts for different categories based on time and territoriality and are known as Google's *Year in Search* (formerly: *Zeitgeist*). They are presented by Google as a simple algorithmic mirror of aggregated human interest. However, Gillespie (2014) argues that 'algorithms designed to offer relevant knowledge also offer ways of knowing.'

What 'ways of knowing' are offered by annual search query metrics, and how are they accepted, ignored, perpetuated, or questioned by journalists of traditional media outlets? This paper answers these questions based on a qualitative study of national media and major newspapers reporting on Google's *Year in Search Switzerland* over the last years. It specifically focuses on how the six political dimensions of public relevance algorithms (Gillespie, 2014) have been addressed. Based on Switzerland's status as an officially multilingual country the paper also illustrates the marginalization of linguistic minorities as territoriality-based quantitative search query metrics cater to the majority by constructing, algorithmically, a national public of search users that does not exist as such.

Introduction

Web search engines provide various search query metrics in different formats for diverse audiences. These metrics include calculated 'related searches' for users, statistics about specific queries for advertisers, and annual 'trends' by country that are often reported on by traditional media outlets. These annual statistics present top-ranking query charts for different categories based on time and territoriality. The most well-known example is Google's *Year in Search* (formerly known as *Zeitgeist*). Google has actively encouraged reporting on its annual search query statistics through press releases and additional dedicated webpages for the press, notably in 2007, 2008, and 2009. In its corporate blog, Google defines *Zeitgeist/Year in Search* as 'How The World Searched' (2011), the 'Collective Consciousness' (2009), or 'Moments That Defined 2014.' The top-ranking query charts are presented as a neutral algorithmic mirror of aggregated human interest.

Suggested Citation (APA): Jobin, A. (2015, October 21-24). *An imaginary algorithmic public: how media report on search query metrics.* Paper presented at Internet Research 16: The 16th Annual Meeting of the Association of Internet Researchers. Phoenix, AZ, USA: AoIR. Retrieved from http://spir.aoir.org.

The political dimensions of algorithms

Unpacking the notion of 'platform' and the ramification of its use, Gillespie (2010) reminds us that vocabulary and discourses by corporate stakeholders about their own technologies are not innocent. By framing *Year In Search* as a neutral algorithmic mirror of search users' common interests, Google refuses to acknowledge that data does not speak for itself (cf. boyd & Crawford, 2012, p. 666; Couldry, 2014) and erases its own contributions. It insinuates that these algorithmic charts are indeed ranking our interests, our 'collective consciousness', our 'defining moments'. Google's *Year In Search* shows how 'algorithms designed to offer relevant knowledge also offer ways of knowing' (Gillespie, 2014): our interests are, supposedly, reflected in the very word we type in Google's search query field and can be aggregated, quantified and ranked unambiguously. This is the narrative inherent in Google's *Year In Search*.

However, algorithms are not neutral: they are social objects constituted by and simultaneously co-constituting values, norms, and knowledge, whether they are understood as a technology (cf. Bijker, Hughes, & Pinch, 1987; Flanagin, Flanagin, & Flanagin, 2010; Fulk, 1993), as a procedure – cf. Rieder (2012) and Cardon (2013) for an analysis of social values encoded in PageRank – or as complex sociotechnical artifacts (Gillespie, 2014; Pasquale, 2015). According to Gillespie, there are 'six dimensions of public relevance algorithms that have political valence' (Gillespie, 2014). These political dimensions range from logics of classification to narratives by and about the algorithms themselves.

Diakopoulos argues that due to the complexity of algorithms 'journalists will [...] be needed to frame, contextualize, and explain the transparency information about algorithm' to the public (Diakopoulos, 2014, p. 29). Since Google's *Year In Search* finds its way into traditional media reporting, this paper analyses how traditional media frame Google's algorithmic top-ranking query charts and whether these six political dimensions of algorithms are addressed as such by journalists. Are the 'ways of knowing' offered by annual search query metrics, perpetuated, ignored, or questioned by journalists of traditional media outlets?

Reporting on an imaginary national public

In order to investigate whether articles of traditional media accept Google's discourse of *Year in Search* as a neutral algorithmic representation of aggregated human interest I have undertaken a qualitative analysis of written articles by five[1] different Swiss media outlets (cf. Fig 1 below). Two types of media have been selected: national websites by public broadcast institutions (RTS and SRF) and major quality newspapers (Le Temps, 24heures, NZZ, Tages-Anzeiger).

1 Six media are listed, but according to its online archive, no article was published by Le Temps on the topic during the investigated time period, a hypothesis confirmed on request by the journalist of Le Temps who had signed a previous article on Google Zeitgeist.

Media outlet	Language	2011	2012	2013	2014
RTS (public national TV/radio)	French	I	I*	I	II
Le Temps (daily newspaper)	French	-	-	-	-
24heures (daily newspaper)	French	-	I	I	I
SRF (public national TV/radio)	German	n.a.**	I	II	I
NZZ (daily newspaper)	German	I	I	II	I
Tages-Anzeiger (newspaper)	German	I	I	I	I

* Two more news agency articles were not specifically about Switzerland.
** No available online archive before 2012

Fig. 1: Number of articles on Google's *Year in Search Switzerland* per media per year

Both French and German speaking media have been considered, because Switzerland has a particular, multilingual status: although there is a two-third majority of German speaking Swiss inhabitants, the three other languages spoken in particular geographical segments of the country (French, Italian and Rhaeto-Romanic) are also official national languages, with French being the most widespread among the three. These particular geographical segments are not naturally delineated by any political boundary but constitute linguistic and cultural boundaries in and by themselves (Blum & Prinzing, 2013; Lüdi, 2008; Widmer, 2004). The articles have been collected by accessing the media websites and/or online archives.

Based on Gillespie's 'six dimensions of public relevance algorithms that have political valence' (2014) I have identified, for each dimension, an operational example that has been mentioned at least once in reporting on Google's *Year in Search Switzerland* (cf. Fig. 2 below). The articles have then been analyzed thematically, whether they mention each of the six political dimension with regard to the annual top-ranking query charts.

Political dimension	Operational example
Patterns of inclusion	'Adult' queries have been excluded from the rankings
Cycles of anticipation	Various functions such as autofill, search-as-you-type, 'did you mean' etc. invite search users to choose existing and statistically probable search queries
Evaluation of relevance	Some charts are based on 'trends' (i.e. increase of search volume) while others are based on absolute volume
Promise of algorithmic objectivity	The data is provided by a for-profit corporation, which has selected the categories and implemented the rankings
Entanglement with practice	A search query is not always a manifestation of interest. (e.g.: navigational queries)
Production of calculated publics	There is no 'Swiss' public but a linguistic majority as well as several linguistic minorities.

Fig. 2: Operationalization of the six political dimensions of public relevance algorithms (Gillespie, 2014)

Unsurprisingly, Google's quantitative search query metrics eclipse Switzerland's linguistic minorities. Google's *Year in Search Switzerland* lists 'Swiss celebs' and 'Swiss sportspeople', but it would be more accurate to outright name the categories 'Swiss-German celebs' and 'Swiss-German sportspeople'. Indeed, only very few queries listed refer to the French-speaking part of Switzerland, and none to the Italian- or Rhaeto-Romanic-speaking part. By basing its top-ranking charts on a quantitative majority and labeling it 'Year in Search Switzerland' Google ignores one third of Switzerland's population and creates the image of a Swiss public of search users that does not exist as such. In Switzerland, the media of a linguistic minority are more aware of the linguistic majority than the other way round (Blum & Prinzing, 2013). This is reflected by the sampled articles: Swiss French media are more likely to point out Swiss German specificities on the top-ranking query charts than are Swiss German media to notice the absence of other linguistic references. On the contrary, Swiss German media mostly repeat Google's minority-blind discourse of Swiss people's supposed interests. While this is made visible through Switzerland's specific linguistic situation, the underlying algorithmic logic is in no way particular to Google's *Year in Search Switzerland* only.

The epistemological focus on how media report on the political dimensions of these algorithms helps reveal 'the *gaps and breaks* [sic] in our languages of social interpretation, authorized by the myth of big data, on which we must focus' (Couldry, 2014, p. 889). Many of the sampled articles, for instance, fail to realize the difference between Google's *Year in Search* charts listing increased search volume and those listing absolute search volume. As a consequence, some articles even assert misinterpretations of comparative search query frequencies.

In addition, Google's categories of the top-ranking query charts are rarely called into question, although all articles except one do mention Google as the data provider. And whereas one newspaper points out the exclusion of 'adult' search queries each year, this pre-selection of qualifying search queries is mentioned only once by another media outlet. Last but not least, almost all media rehash, uncritically, Google's equivalence of *s*earch as a manifestation of interest. Overall, except in one of the newspapers, many of the political dimensions of the algorithmic search query charts remain unaddressed most of the time.

The paper suggests that, by not adequately addressing the political dimensions of the algorithmic search query charts, traditional media reinforce the performativity of these top-ranking charts all while contributing to stabilize Google's narrative that these charts are simply a neutral algorithmic mirror of people's interest.

References

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. London: The MIT Press.

Blum, P. D. R., & Prinzing, P. D. M. (2013). Medien und sprachliche Segmentierung. In K. Imhof, R. Blum, H. Bonfadelli, & O. Jarren (Eds.), *Stratifizierte und segmentierte Öffentlichkeit* (pp. 149–166). Springer Fachmedien Wiesbaden.

boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679.

Cardon, D. (2013). Dans l'esprit du PageRank. *Réseaux*, *177*(1), 63–95

Couldry, N. (2014). Inaugural: A necessary disenchantment: myth, agency and injustice in a digital world. *The Sociological Review*, *62*(4), 880–897.

Diakopoulos, N. (2014). *Algorithmic Accountability Reporting: On the Investigation of Black Boxes.* New York: Tow Center for Digital Journalism, Columbia University. Accessible at: http://towcenter.org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/

Flanagin, A. J., Flanagin, C., & Flanagin, J. (2010). Technical code and the social construction of the internet. *New Media & Society*, *12*(2), 179–196.

Fulk, J. (1993). Social Construction of Communication Technology. *Academy of Management Journal*, *36*(5), 921–950. http://doi.org/10.2307/256641

Gillespie, T. (2010). The politics of "platforms." *New Media & Society*, *12*(3), 347–364.

Gillespie, T. (2014). The Relevance of Algorithms. In *Media Technologies*. Cambridge, Massachusetts: Mit University Press Group Ltd.

Google. (2009). Zeitgeist 2009: the collective consciousness. Retrieved from http://googleblog.blogspot.com/2009/12/zeitgeist-2009-collective-consciousness.html

Google. (2011). Zeitgeist 2011: How the world searched. Retrieved from http://googleblog.blogspot.com/2011/12/zeitgeist-2011-how-world-searched.html

Google. (2014). A Year in Search: the moments that defined 2014. Retrieved from http://googleblog.blogspot.com/2014/12/a-year-in-search-moments-that-defined.html

Lüdi, G. (Ed.). (2008). *Sprachenvielfalt und Kulturfrieden: Sprachminderheit - Einsprachigkeit - Mehrsprachigkeit: Probleme und Chancen sprachlicher Vielfalt*. W. Kohlhammer Verlag.

Pasquale, F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press.

Rieder, B. (2012). What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index. *Computational Culture*, (2). Retrieved from http://computationalculture.net/article/what_is_in_pagerank

Widmer, J. (2004). *Langues nationales et identités collectives: l'exemple de la Suisse*. Paris: Harmattan.