User Searches on Multiple Channels during the US Presidential Elections 2020

Christoph Glauser[1*], Jacques Savoy[2], Loris Schmid[1,2]

[1]IFAAR, Institute for Applied Argumentation Research, Marktgasse 55, 3011 Bern, Switzerland;[2]Computer Science Department, University of Neuchatel, Rue Emile Argand 11, 2000 Neuchatel, Switzerland

ABSTRACT

The present paper is taking up the actual discussion about how digital channels are having an impact on democratic decision making processes in modern election campaigns. Systematic empirical monitorings of user search behaviors during the US Presidential Elections in 2016 and 2020 on multiple channels show first evidence on how search volumes are distributed over the online channels and how these searches have influenced the campaigns of Donald Trump and Joe Biden. The Search engines Bing and Google dominated. However, high search volumes were also observed on the Chinese search engine Baidu. Altavista-Yahoo and Twitter as well as social networks like Facebook, Wikipedia and even e-shops follow on the spot. The data show how Donald Trump might even have created an "information overload" when he announced his positive Corona test result late in the campaign.

Keywords: Presidential elections; User monitoring; Search volumes; Multi-channel; Tech intermediates; Digital; Online; Campaign analysis

INTRODUCTION

The aim of this research project is to monitor as many online channels (tech intermediates) in the US as possible, including Search Engines (SE), Social Media (SM) and E-Shops (ES), in order to examine the user interests based on their active searches on the internet during the Presidential elections in 2020 [1]. Since the introduction of the html format by Tim Berners-Lee in 1993, early SEs' like Altavista, followed by Yahoo and many other, have been publically used to find the most relevant web pages on the web in response to user search requests (page rank). Today SM networks are functioning as "SE's" as well, even if sometimes in a closed user group context. To better understand users' needs on SE's, SM and even ES it is necessary to analyse the frequency distribution of actively submitted search queries on as many portals as possible, which is what the present study is doing. Tech companies usually use the gathered data for ranking information and for commercial advertising purposes. In our research they are systematically monitored for empirical and scientific analyses of election campaigns.

After scientists of the University of Neuchâtel in Switzerland have already analyzed the 2016 by Presidential elections, the scope was extended to more tech intermediates in 2020 [2]. In the SE-research community, research about search volumes is still primarily led by rather narrow Google AdWords search data on Google trends, as most of the recent studies show [3]. Other research initiatives focus entirely on single channels like Twitter but usually rather on the content part as opposed to the search volume of Twitter as a SE. Therefore the tech intermediates under consideration in the presented project have

been systematically extended to include more sources such as Google, Altavista Yahoo, Bing, Wikipedia or facebook, Youtube, Twitter, Instagram and eBay, Amazon, Alibaba etc. Up to 14'103 channels worldwide were anaylsed in order to extend the discussion about the "impact" of digital secondary sources. This is important to understand the decision making process of users and voters.

By one of our earlier hypothesis, user requests with a proper name are potentially indicating the popularity of the respective presidential runner. Important to note is that the authors do not claim a "stimulus response" model reaction as originally defined by Maletzke in the early 60s and many other mass media scientists ever since have published: In short, this model claims that information which is released by a sender and getting media coverage directly leads to an ultimate reaction of the recipient [4].

In our online case this would mean that the candidate with the highest media coverage would automatically be the one that's being elected. As an example: "If Donald Trump publishes more frequently to vote for him (as compared to Joe Biden) and consequently more citizens would go to vote for him". Our data on the digital channels from former election campaigns show at least one strong correlation: The more intensive a campaign is getting, the more search volume is generated on the user side.

In 2016 the significantly higher search volumes of Donald Trump were interpreted as an early indicator for his victory in 2016. Previous publications of the UMUSE project of mid-September 2016 already showed Tump far ahead on the SE's as compared to Hillary Clinton [5]. In 2020 Trump was again five times leading on the search volume

during the campaign, but it seems as if he has somehow "overdone", especially when he was using his positive Corona test result for campaigning. This caused campaign effects, which have never been measured before to this extent on the web, with more than 100'000 million searches on a 30 day's average on all the US digital channels. These effects of Donald Trump's media "information overload" might well have led to Joe Biden's victory, as American voters became suspicious after the rather quick recovery of the president shortly after his announcement of being affected himself by the virus. These are effects which yet will have to be examined in further research projects on Donald Trump's Corona "hype" in the middle of the "hot phase" of the 2020 election campaign [6].

In the present Swiss computer and mass media science research UMUSE project the following hypotheses needed to be verified:

1. The first hypothesis goes beyond digital attention and aggregated search volumes: It claims that user interest in modern election campaigns in the digital age are much more distributed over various tech intermediaries (channels), than expected.

2. The second hypothesis claims: Users search for candidates and for generic topics on those channels which they predominantly use, no matter if it is a classical search engine, social media or even an e-shop platform.

3. The third hypothesis: The search volume may change significantly depending on the issue, the candidate or even the particular interest of user groups (economical, political, at work, at home, foreigners and health care in the US etc.).

METHODS AND MOTIVATION

Thanks to advances in API and mainly Java programming technology, the interfaces for thoroughly analyzing search volumes on multiple digital channels have been continuously improved during the last ten years in close collaboration with computer science departments of the Universities of Neuchâtel, Berne and Freiburg in Switzerland, where students have programmed various API's and software tools in order to collect search volume results on multiple digital platforms for scientific and analytic purposes.

Between March 2020 and November 2020, during the entire US presidential election campaign 54 tech intermediaries were monitored in the US context. By employing Application Program Interfaces (API), specially programmed software classes have been connected to all relevant platforms, using search frequency or keyword related statistical data, offered by the providers themselves, as well as used by the users directly. This computer science project collected full scale relevant data for the US digital "universe" in order to provide a digital multi-channel overview for the whole US Presidential election campaign.

Search volume data is extracted from 14,103 search engines, social networks and other electronic platforms such as e-shops and media on a multinational level. Each of these sources has its designated API and hence needs to be handled separately. Moreover, the data format may change without prior notification which requires ongoing daily monitoring of the integrity of the empirical results acquired from that source. The harvested data are then analyzed systematically to reveal and validate the queries submitted by Internet users and the frequencies of those inquiries.

Some additional parameters need to be specified. First, the data crawling can be parameterized to take in account only a predefined geographical country such as the US. In 2016 the comparison was done also with comparative search volumes with a sample of Canada and UK. In the present study, the targeted users are only located in the United States. Second, it is important to examine and compare numerous SEs', Social Networks (SMs'), and E-Shops (ESs') to assure a high level of country and online user coverage. Focusing only on a single (or a few) service(s) would run the risk of obtaining biased or inaccurate data. Third, the gathered empirical data were validated and tested for reliability. Therefore, patterns are identified and topics are reasonably compared across most of the popular tech intermediaries. The best scope for measuring valid search volume data on various digital channels is the ∑ of the arithmetic average of search scores over the previous 30 days at any given time. Most of the sources worldwide usually offer a 30 day data range, some more, some less. The raw results have been measured as "moving averages" of these 30 days for each day, for each candidate or keyword of the campaign and on each channel. Reliability testing is done by comparing the multi-channel results of all search queries, with the entire number of active users in a given country (by top level domain). In the case of the US for this universe of active users, it is based on the ITU figures which identify in the year 2020 approximately 284 million active users in the US.

The innovative finding technology (developed and improved between 2010 and 2020) allows scientists to examine, monitor, and systematically compare the evolution of the candidates and issues digital demand throughout the entire US electoral time span. In prior work, some preliminary studies have been conducted on a smaller scale when analyzing campaigns or online behaviour about Covid-19 in Switzerland [7]. To observe the evolution of each candidate's popularity, their names were used as queries (namely "Donald Trump," "Joe Biden"). Then, the data were gathered daily throughout the entire period of the election campaign from February 1st, 2020 until November 20th, 2020. Due to intermittent technical issues, there are no data available for only a few days during this whole period. We decided to use the full name of each candidate as a query term to avoid possible false matches with other public figures carrying the same surname (e.g. Ivanka Trump) or corresponding to a name related to different named entities (e.g. Trump Tower).

RESULTS AND DISCUSSION

Multi-channel results

Showing the widely spread distribution of the active searches over various digital tech intermediaries, the first hypothesis i) could be verified with the measured API-results about both candidates. The results show that apart from the "big" channels like Google, facebook, Twitter and Amazon there are also other equally or even more important digital channels which are intensively used by disperse internet users during an election campaign.

Figure 1 is showing the distribution of search volumes for the keywords "Donald Trump" and "Joe Biden" between March 1st and November 20th on the most frequently used 18 digital channels (n=54) in the US. The data is showing that most searches in the long term took place on Bing Mirosoft SE, indicating that most of the 288 million active users in the USA have probably searched in some Microsoft environment (often at work) using Bing search engine for "Donald Trump" and for "Joe Biden" [8]. Bing is rather closely followed by the searches on the Google SE. The result indicates a particularly high interest in both candidates by people working in an office, delivering services or just using various digital business devices and connections during the whole campaign.
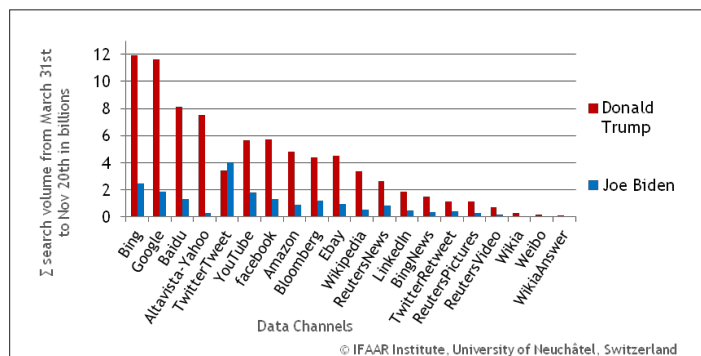
Figure 1: The graphic shows the aggregated search volume by the 20 most frequently used tech intermediate channels in the USA for the two main candidates during the US Presidential election campaign from March 1st, 2020 until November 20th, 2020.

Displayed in Figure 2 are the ongoing searches on Bing. They show how the campaign searches have slowly taken up rather late starting mid-August and how they have increased continuously for both candidates towards the date of the final elections on November 3rd. Figure 3 compares Google and Altavista-Yahoo and the candidates in 2016 and 2020.
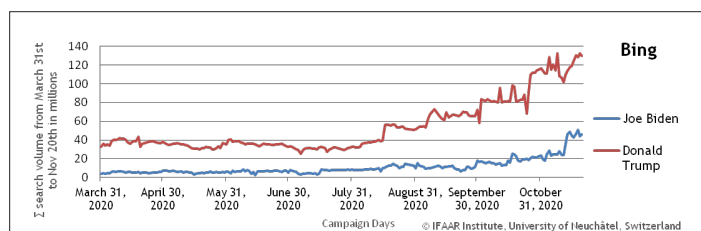

Figure 2: The graph illustrates how moving averages of search volume on Bing has evolved during the campaign for Joe Biden and Donald Trump searches.
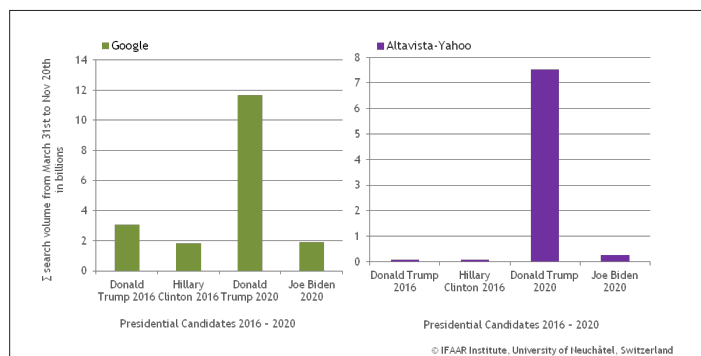

Figure 3: Aggregated search volumes for the main candidates in two comparable examples Google and Altavista-Yahoo for the campaigns 2016 and 2020.

If Google would add YouTube searches statistically to Google, it would rank number one instead of Bing. But it was important in this project to treat all the channels delivering search volumes fully independently. Third ranks surprisingly the popular Chinese search engine Baidu with still 8 billion searches during the whole period of the elections, indicating that probably a lot of Chinese users within the US have been searching rather actively about "Donald Trump" assuming interest there for several measures and verbal attacks against Chinese trade and towards the end of the campaign also against Tiktok by president Trump. The searches for Joe Biden on Baidu during the whole campaign have been significantly lower than the ones for Donald Trump.

Still significantly used in the USA is the search engine Altaviste-Yahoo, which is ranking fourth. The search volume top list number five, Twitter, is followed by YouTube and Facebook as number six and seven. Amazon, Bloomberg and Ebay can already be considered as being "particular interest" channels, so do the searches on Reuters News for example. The comparably high volume of search on these e-shops and business channels supports hypothesis ii) because they show that generic interest for political candidates' takes place on channels which have so far rather been associated with other purposes, like shopping or like particular economical or stock market interests. Obviously online users simply don't care and happen to search by accident, just on the platform they already are, when it comes to the active search about one of the candidates in the election campaign. This may give an alternative view on the campaigning. The "entry point" for the candidates varies considerably as our data shows. This does not mean that users might not finally end up on a candidates' campaign website or promotion at the end of their search, but it shows they intuitively start wherever they are in the very moment, launching an ultimate search query like "Donald Trump" or "Joe Biden".

Wikipedia, which is also becoming more and more like a search engine, ranks between Ebay and ReutersNews. Twitter on rank five, would even be the top rank for Joe Biden, however, he only increased on Twitter towards the far end of his campaign when the searches for Donald Trump on Twitter already had partially declined, except for his Corona announcement period:

Although Twitter is considered to be a special interest channel for PR, journalists and media and for some close community followers, it has become one of the most important sources of information on the elections throughout the year 2020 again like in 2016, strongly dominated of course by Donald Trump. But towards the very end of the campaign the search volumes for Joe Biden on Twitter have significantly increased in comparison to Donald Trump. Earlier in the year Trump presented a more visible profile sending more than 43 tweets per day (September) and having more than 87 million followers. Biden writes around 11 tweets per day (September) for around 10.7 million followers. When looking at the stylistic aspects, Trump re-tweeted much more (35% of his tweets) as Twitter studies by Jacques Savoy show [9]. But Twitter searches indicate that the journalists have started to make more inquiries (searches) on Joe Biden staring towards the end of September as the chart above is showing well.

Biden beats clinton, Trump has "overloaded"

The US election campaign of 2020 was historically unique event. The pandemic situation in the US made it fairly impossible to run a "normal" election campaign for both parties and candidates.

These particular circumstances had the effect that secondary online media became much more important channels. Technical intermediates were in place to carry on the political messages of the campaign to the potential voters whereas direct "live" contacts were extremely limited due to the pandemic situation. All electronic media, in particular TV became an exceptionally important source even more than it has been in previous campaigns. In conclusion, in 2020 the online channels became extremely important technical intermediates between political parties, candidates and voters out in the public sphere (Figure 4).
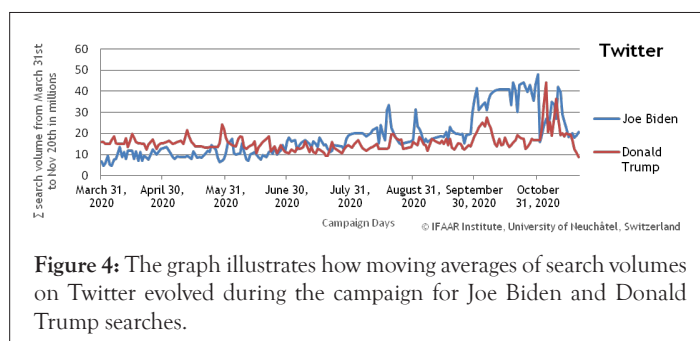
Figure 4: The graph illustrates how moving averages of search volumes on Twitter evolved during the campaign for Joe Biden and Donald Trump searches.

Comparing entire search volumes from the 2016 to the 2020 campaign shows that the search for "Donald Trump" almost doubled. Joe Biden managed to get more searches but only towards the very end of the campaign (Figure 5). If, however, predictions could have been made based on the search volumes, it was only by mid-October, when the searches for Joe Biden suddenly came up to an unexpected high level.
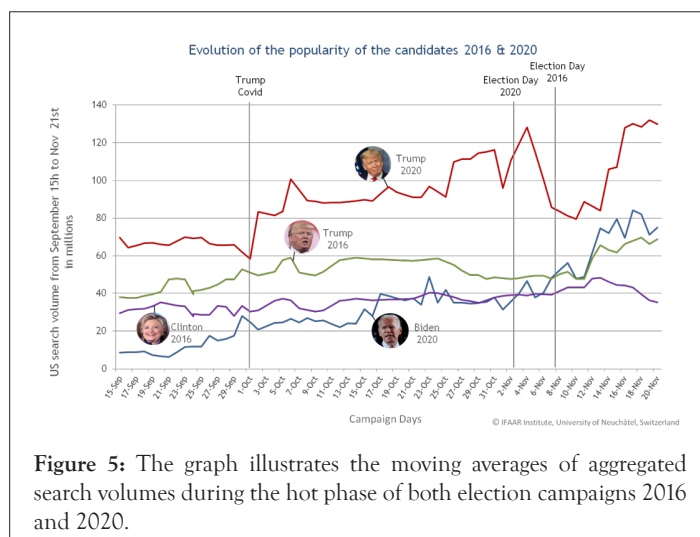

Figure 5: The graph illustrates the moving averages of aggregated search volumes during the hot phase of both election campaigns 2016 and 2020.

Donald Trump managed to reach an all-time high search volume of 100 million searches in the 30 day moving average after October 3rd, when he disclosed that he was tested Corona positive, after having almost completely ignored COVIDs' pandemic danger for several months. This indicates the dipping point of his campaign, where Donald Trump looked more and more like a desperate "phishing for attention" candidate to average American voters. On the other hand Joe Biden with his long lasting silence throughout the campaign year could obviously steadily mobilize diverse voters like African Americans, Latinos and women towards the end of the hot race. He even managed to beat Hillary Clintons' scores in 2016 towards the end of the campaign. The data supports the third hypothesis iii): The search volume may change significantly depending on the issue, the candidate or even the particular interest of user groups (economical, political, at work, at home, foreigners and health care in the US etc.). As the Corona example shows, not only offline or electronic media campaigns can "overdo". Online users who are also voters are getting suspicious if the coverage exceeds a certain level of "advertising pressure". This can even lead to staying at home at the Election Day [10]. The Corona topic on the search volume shows that this issue may have caused the "information overload" in Donald Trump's campaign, because he was no longer taken seriously even by a part of his own republican voters.

CONCLUSION

Tech intermediates are changing the way political campaigns take place. User interests in election campaigns in the digital age are much more distributed over various tech intermediates (various and variable online channels) than expected. Search volumes, if not limited to certain channels can be a strong indicator for the popularity of candidates and topics in modern election campaigns as well as for the trends.

Users usually search for candidates and for generic campaign issues on those channels, which they using preferably, no matter if it is a classical search engine, a social media or even an E-Shop platform. The example of "Corona" shows that search volumes can change significantly during a campaign depending on the issue, the candidates or even the particular interest of user groups (economical, political, at work, at home, foreigners, health care in the US etc.). If a candidate "overloads" during a campaign, this can even lead to voters staying at home on the Election Day.

More research in the Mass media, political, as well as computer sciences is needed to understand how search volume data can be applied in future election campaigns to make better predictions, or at least to identify more solid trends even in real time. Tech intermediates are invited to offer more reliable data to science by API in order to better understand how search is influencing political campaigns. Closing down profiles like Twitter and Facebook did, is probably not a smart way to achieve more transparency. Hopefully there will be more sophisticated approaches to address many of these new media topics in 2024.

ACKNOWLEDGEMENT

REFERENCES

1. Oehmer F, Udris L, Eisenegger M. Zur Wirkung von Tech-Informations-Intermediären auf die (Schweizer) Öffentlichkeit – Chancen und Risiken. fög. Zürich 2020.

2. Savoy J. Trump's,Clinton's Style and Rhetoric During the 2016 Presidential Election. Journal of Quantitative Linguistics. 2018; 25(2): 168–189.

3. Wild J, Chen W, Lohmann S. COVID-19 and the future of US fertility:What can we learn from google? IZA, Bonn (2020).

4. Maletzke G. Psychologie der Massenkommunikation. Hamburg.1963;10(2):185-187.

5. Glauser C, Savoy J, Gasparian M, Nada N. UMUSE: Monitoring of the US Election.

6. Proceedings SVPW, St.Gallen (2017). Informationsüberflutung in Zeiten der Digital Guide IONOS.

7. C. Glauser, L. Schmid, Kurpicz-Briki M. Identifying the public interest in covid-19 contact tracing apps in switzerland based on online search behaviour in: Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance. 2020;120-123.

8. Committed to connecting the world,ITU.

09. Glauser C. Schmid L, Savoy J. UMUSE: User Monitoring of the US Presidential Election. In: Proceedings ICICT 2021, to appear. Springer, London.

10. Bernard L. Fraga, Zachary P, James S.New Georgia runoffs data finds that more Black voters than usual came out. Trump voters stayed home.Washington Post.