MONITORING OF THE LAST US PRESIDENTIAL ELECTIONS

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

[1]Department of [2]IFAAR Institute, Bern, Switzerland
[2]Computer Science Dept., University of Neuchatel, Switzerland

ABSTRACT

This paper presents the results of a new monitoring project of the US presidential elections with the aim of establishing computer-based tools to track in real time the popularity or awareness of candidates. The designed and developed innovative methods allow us to extract the frequency of queries sent to numerous search engines by US Internet users. Based on these data, this paper demonstrates that Trump was more frequently searched than the Democratic candidates, either Hillary Clinton in 2016 or Joe Biden in 2020. When analyzing the topics, it is observed that in 2020 the US users had shown a remarkable interest in two subjects, namely, Coronavirus and Jobs (unemployment). Interest for other topics such as Education or Healthcare were less pronounced while issues such as Immigration were given even less attention by users. Finally, some "flame" topics such as Black Lives Matter (2020) and Gun Control (2016) appear to be very popular for a few weeks before returning to a low level of interest. When analyzing tweets sent by candidates during the 2020 campaign, one can observe that Trump was focused mainly on Jobs and on Riots, announcing what would happen if Democrats took power. To these negative ads, Biden answered by putting forward moral values (e.g., love, honesty) and political symbols (e.g., democracy, rights) and by underlying the failure of the current administration in resolving the pandemic situation.

KEYWORDS

E-government, Online searches, Search engines, Discourse analysis, Political science
.

1. INTRODUCTION

Today's computational technology offers new perspectives in social sciences for studying a large volume of data in various formats such as text, images and video [1]. The Internet can be used as a source to detect early trends in an on-going event such as an election, vote, social movement, or as a source for opinion monitoring. It could also help prevent the misinterpretation of results from polls during a campaign as demonstrated by the erroneous predictions during the Brexit and 2016 US presidential campaign.

In this perspective, the main objective of our project has been to analyze the 2020 US presidential election with some comparisons with the 2016 campaign. Our real-time research aimed at combining various scientific methods simultaneously for analyzing the US elections to create a more differentiated view of what the citizens want and what the candidates talk about. The core hypothesis is that the frequencies of the queries sent to search engines reflect the people's active interest in these topics (e.g., "Donald Trump," "Joe Biden," "gun reform," and "illegal immigration"). Therefore, the higher the frequency of a given query, the more popular or interesting it is. Moreover, when we are not interested in a topic, we will not search for it. Such a hypothesis is related to various web-based systems such as the PageRank algorithm implemented by Google to define the popularity of each webpage in which the higher the number of incoming links is a usefulness indicator of the targeted page [2].

The rest of this paper is organized as follows. The next section presents an overview on related research in computer-based analysis in the political domain. The third section presents the technology developed and its application to the candidates' names. The fourth describes and applies a monitoring system to a set of selected topics reflecting the main concerns of the digitally active US citizens. The fifth section analyzes the candidates' tweets to reveal their rhetoric and their main interests during the 2020 campaign. A conclusion draws the main findings of this study.

2. RELATED WORK

The political domain has been the subject of different computer-based studies, mainly to analyze written messages. The more intensively studied aspects have been governmental speeches in which the differences could be explained by the presence of distinct political parties. Labbé *et al.* [3], [4], [5] analyzed the Canadian, Quebec, and French governmental addresses and showed that the differences between political parties tend to disappear once they are in command. The time period represents an essential factor in explaining the variations between presidents or prime ministers. As other important reasons explaining a real change in theme, style and rhetoric, one can mention the presence of a strong leader or dramatic events (world war, The Great Depression) [6], [7]. Focusing on the US, other studies confirm these conclusions, as for example, based on the *State of the Union* [8] or the inaugural addresses [9].

Each political party aims for an utmost online presence and wants to be viewed as distinct from its opponents. The distinctions between political parties can usually be detected via textual analyses of their delivered content. Based on social networks, Sylwester & Purver [10] have shown that differences tend to be correlated with psychological factors. For example, statistically-speaking, words conveying positive emotions tend to occur more frequently in Democrats' tweets than in Republican ones, so do swear words, and first-person pronouns (e.g., I, me). For Raubach [11], positive emotion words should be associated to the party in power and not simply attached to a given party. Focusing on legislative addresses, Yu [12] demonstrated that machine learning methods can be trained to classify congressional speeches according to political parties. In another study, Yu [13] revealed that political feminine figures tend to use emotion words more frequently, and use more personal pronouns than men [14].

Other studies aim to analyze people's consent and approbation of government actions. In this perspective, O'Connor *et al.* [15] have demonstrated the ability to estimate the popularity of the Obama administration based on tweets. This study found a positive correlation between the presidential approval polls and positive tweets containing the hashtag #obama. Such a selection strategy, however, produces a low recall since many tweets about Obama's administration are overlooked if they don't contain the aforementioned hashtag. As a tweet is rather short (140 characters, now 280, equivalent to an average of eleven / twenty words), the sentiment estimation is simply the count of the number of positive and negative words based on the OpinionFinder dictionary [16]. Mohammad *et al.* [17] have analyzed sentiment in tweets during the 2012 US presidential election. They found that *disgust* was the most frequent emotion followed by *trust*. Moreover, in the harvested tweets, they found that negative emotions occurred twice as much as positive ones did.

The electoral process tends to exacerbate political differences and exhibits more emotions [18]. To detect and measure them, Young & Soroka [19] suggested a strategy based on several dictionaries. The suggested approach is rather similar to O'Connor *et al.*'s work [15], counting the frequency of occurrence of words appearing in a dictionary of positive or negative emotion words. Using the DICTION system [20], Bligh *et al.* [21] analyzed the rhetoric of H. Clinton during the 2008 presidential election. Clinton appears more feminine than the other candidates, using more *I* than

we, and showing a higher frequency of words belonging to the class "human interest" (e.g., family, man, person, etc.). Recently, Hart [22] analyzed Trump's rhetoric and concluded that Trump is the president of the extremes, presenting either a high or low level depending on the target emotion or rhetorical concept.

As another point of view, a set of studies outlined the psychological portrait of political figures. Based on the LIWC system [23], Slatcher *et al.* [24] determined the personalities of several political candidates during the 2004 US presidential election. They defined the psychological portrait based on both single measurements (e.g., the relative frequency of various pronouns, positive emotions, etc.) and a set of composite indices reflecting the cognitive complexity, presidentiality and perceived honesty of each runner. These personality measurements were in agreement with different opinion polls. For example, G.W. Bush tends to frequently use the pronoun *I*, positive emotion words (e.g., happy, truly, win), and the future tense. The public perceives J. Kerry as somewhat gloomy, serious, and cold, as he tends to adopt negative emotion expressions (e.g., sad, worthless, cut, lost) and physical words (e.g., head, ache, sleep).

Finally, various predictive models have been proposed to estimate the outcome of an election or a vote. The conventional model is based on a nationwide poll, however, other models also exist including ones specific to US presidential elections [25] as well as social networks-oriented ones [26]. Studies tend to suggest that social media are too biased to be considered reliable sources of information [27].

As a brief conclusion, one can deduce that computer-based analysis in the political domain generally focuses on the governmental level and mainly on existing text sources such as official speeches, party manifestos and, more recently, tweets. The electoral process [28] and how it is influenced, on the other hand, is less studied and the evolution of both candidates' popularity and demand of the public is analyzed through poll estimates or user surveys. This study presents another approach to obtain similar information through the analysis of the active demand on search engines and social media.

3. FOLLOW-UP ON THE POPULARITY OF THE CANDIDATES

Thanks to recent advances in natural language processing [29] and API-technology, we are able to extract data from more than 1,457 search engines (e.g., Google, Bing, Yahoo!, Ask, Baidu, Reuters, Technorati, MetaCrawler, Search.com, etc.) and social network platforms worldwide (e.g., Facebook, Twitter, Wikipedia, etc.). 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 daily monitoring of the integrity of results acquired from each source. The harvested data are then analyzed systematically to reveal the queries submitted by Internet users and the frequencies of those queries. To illustrate these aspects, the Appendix reports the case of two queries sent to seven different sources.

Some additional parameters must be specified. First, the crawl can be parametrized to take account of only a predefined geographical country such as the US, the UK, or Canada. In this study, the targeted users are located in the United States. Second, it is important to examine and compare numerous search engines and social networks to allow a high coverage. Focusing only on a single (or a few) service(s) would run the risk of obtaining biased and/or inaccurate data. Moreover, some sources are occasionally unavailable due to technical and network-related reasons. Third, the gathered data were validated and tested for reliability. Therefore, patterns can be identified and topics can be reasonably compared across most of the popular digital channels. The optimal scope for measuring valid data on various digital channels is the moving average of scores between 3 to 10 days at any given time.

This innovative *finding* technology (developed at the IFAAR institute in Berne between (2010 - 2016) allows us to examine, monitor and systematically compare the evolution of the candidates' digital demand throughout the entire US electoral timespan. In prior work, some preliminary studies have been conducted on a smaller scale when analyzing campaigns or online behavior about Covid-19 in Switzerland [30].

To observe the evolution of each candidate's popularity, their names were used as queries (e.g., "Donald Trump," and "Joe Biden"). Then, the data were gathered daily throughout the entire period of the election campaign as, for example, from March 1st, 2020 until November 3rd, 2020. We decided to use the full name of each candidate as a query to avoid possible false matches with other public figures carrying the same surname (e.g., Ivanka / Melania [Trump]) or corresponding to a name related to different entities (e.g., Trump Tower, Trump impeachment, etc.).

By our hypothesis, user requests with a proper name are potentially indicating the awareness or popularity of the respective presidential runner. We must note that not all submitted queries denote an interest in favor of the corresponding candidate, but the majority thereof are generally assumed to be in favor of the candidate. Having a larger awareness for a given candidate means being the center of interest of the media, appearing more often in the news both in the newspapers and TV shows. To increase his/her media coverage, a candidate could send tweets whose content could hurt the audience or with sentences or expressions not politically correct.

A candidate not getting enough awareness would not win an election. Moreover, we estimate that the proportion of unfavorable queries of a candidate would be similar for all runners. But this aspect is not an essential point. The favorable and unfavorable connotations do not play an important role if awareness is at stake.

On the Web, various tools are using popularity-based information under the assumption that popularity is positively correlated with usefulness. For example, the Google search engine ranking algorithm favors webpages receiving a higher number of incoming links. In this case, a link to a page is viewed as a *vote* in favor of the target page, indicating its relevancy and usefulness [2]. As another example, Cleverdon *et al*. [31] found that only 5% of the bibliographic references in a scientific paper indicate a negative endorsement. References to other entities (authors, webpages, etc.) correspond, in general, to a positive perception of the target entity.
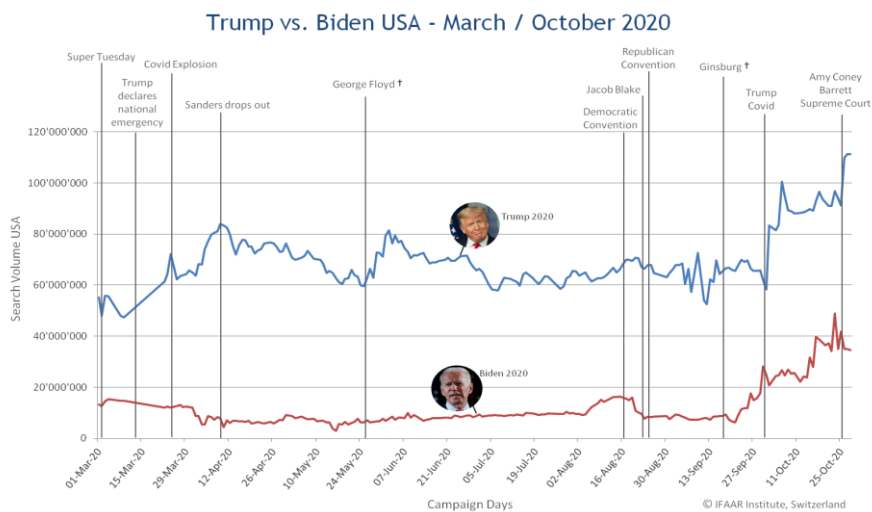

Figure 1. Evolution of the number of queries with the two candidates' names from US Internet users

Based on the harvested data, Figure 1 shows the number of queries sent by US Internet users to search engines and social networks during the US 2020 election. The first point in Figure 1 represents the total number of requests sent March, 1st. The horizontal axis is marked with increments of 20 million requests.

In this graph, vertical lines mark notable events, namely, Super Tuesday (March 3rd), national emergency (March 13th), Sanders' resignation (April 9th), George Floyd's death (May 25th), the Democratic Convention (DNC, August 17th-20th), the Republican Convention (RNC, August 24th-27th), Jacob Blake's shooting (August 23rd), the TV presidential debates (September 29th, and October 22th), death of the Supreme Court judge Ginsburg (September 18th) followed by the confirmation of A. Barrett (October 10th), Trump's infection (October, 1st), and the general election day (November 3rd).

This figure demonstrates that Donald Trump clearly attracts more queries (translating to interest according to our hypothesis) than the Democratic Party nominee. In addition, the difference between both candidates is significant with a mean of 70,527,893 for Trump vs. 11,973,494 for Biden. One can argue that requests for Trump also corresponds to inquiries for the current president. Even when this effect is estimated to 15M queries, Biden was never able to reach the level obtained by the Republican nominee.

A closer look at Figure 1 indicates that one can subdivide this election into three time periods. During the first one, from March 1st to Sanders' resignation (April 9th), the Donald Trump curve increases while that of Biden tends to slowly decline.

The second period goes up to the end of August. One can see a stable but low attention for the Democratic nominee, with a slight increase after Floyd's death (May 25th). Surprisingly, one can note that Biden was not able to enhance his support after Sanders' resignation. For Donald Trump, this interval corresponds to two decreasing trends with a rebound after Floyd's death. This tragic event favored more the incumbent president than his opposition. The two national conventions (DNC, August 17th-20th and RNC, August 24th-27th) were an opportunity for both candidates to boost the interest of US users for themselves.

Clearly the most important time period was the last two months of the campaign. During the first two weeks of September, Biden's popularity was stable but rather low. A clear growth appears after Ginsburg's death (September 18th), a favorable trend also boosted by the first TV debate (September 29th) and the closeness of the election. During the last two weeks, his final level reaches around 40 million requests per day. For Donald Trump, September was not really a good month, with a large variability in the interest of US users for his campaign. The tweet sent on October 1st announcing Trump's infection with Covid-19 produced an incredible boost in the number of queries, reaching more than 100 million requests per day and this trend continued during the last weeks of this electoral campaign. The confirmation by the Senate of the nomination of judge Amy Barrett (October 10th) supports this high attention for the incumbent president.

To obtain a better visualization of this campaign, Figure 2 presents the number of queries obtained by the two main candidates for both the 2016 and 2020 elections. As main findings, one can see that in both cases Donald Trump obtained a higher support. Moreover, during the 2020 election, the president achieved a higher number of requests compared to 2016. For the Democratic nominees, Hillary Clinton depicted a higher backing than Joe Biden but as for the 2020 election, the Democratic nominee was not able to achieve a level higher than Donald Trump.

As for the 2020 election, the main interest for the candidates increased during the last two months. In 2016, Hillary Clinton was closer to Donald Trump than Joe Biden in 2020. This aspect could explain that the polls had announced Clinton's victory for November 8th 2016.
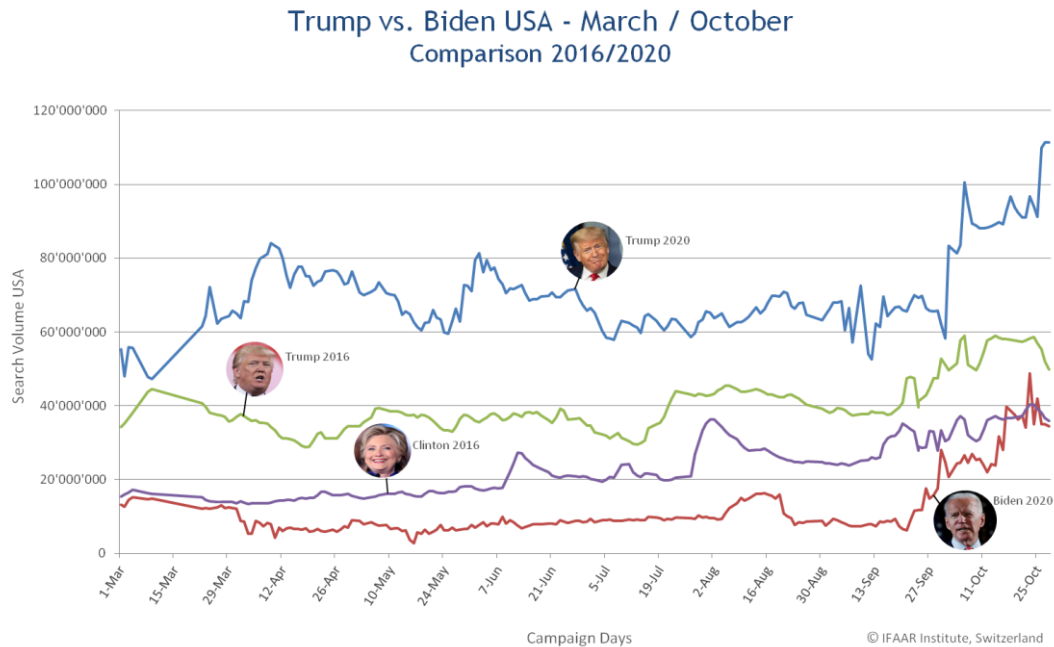

Figure 2. Comparing the evolution of candidates' popularity by search requests in 2016 and in 2020

An election may seem to revolve around a set of candidates and their profiles including personality traits, past actions, balance sheets, or popularity, but it also raises concerns about certain vital matters such as potential changes in healthcare, jobs, education and new legislations. To analyze such aspects, we have used the same finding technology to monitor the people's interest in selected topics as described in the next section.

4. EVOLUTION OF MAIN TOPICS THROUGHOUT THE ELECTIONS

As for each electoral campaign, a selection of topics was picked to represent the most important concerns of the general public. For the 2016 election, Ehrenfreud & Tankersley [32] denoted seven issues which are: immigration, taxes, family, health care, trade, climate change and foreign policy. We kept these subjects for the 2020 election. Of course, this is not a comprehensive list as, for example, the education and gun control topics are missing and have been added to our list. For 2020, we must also include the coronavirus problem and the jobs and unemployment issue.

To analyze the differences between these topics/concerns and to explore their evolution, we have formed a set of search queries. To achieve the most appropriate formulation for each topic, we analyzed the pertinent wording used (and formerly analyzed) in four former US elections (2000, 2004, 2008, and 2012). Moreover, we curated a dictionary with many entries (one entry per topic), for instance, the entry for the topic *Immigration*, is mapped to the queries "illegal immigration," "immigration reform," "immigrant," "Latin immigration," "immigration raids," while the topic (entry) *Healthcare* maps to "Obamacare," "healthcare reform," "medicare," and "health care". A similar strategy has been applied for the other themes. This strategy reflects the fact that a given specific topic can be expressed in numerous forms using different words / expressions [33] without

34

one being predominantly used more frequently than the others. This vocabulary problem is well-known in various applications such as information retrieval [34]. The current solution is however similar to the technique applied by the DICTION [20], [35], [36], or the LIWC tool [23]. Moreover, a given query formulation can cover several meanings, and in our study, with a different intent than the original semantics.

To see the evolution of the US citizens' interests through this 2020 election campaign, Figure 3 clearly depicts the two most important ones, namely Coronavirus and Jobs (or unemployment) from June 1ˢᵗ to November 3ʳᵈ. In addition, this graph presents the topics Education, Immigration, and Black Lives Matter.
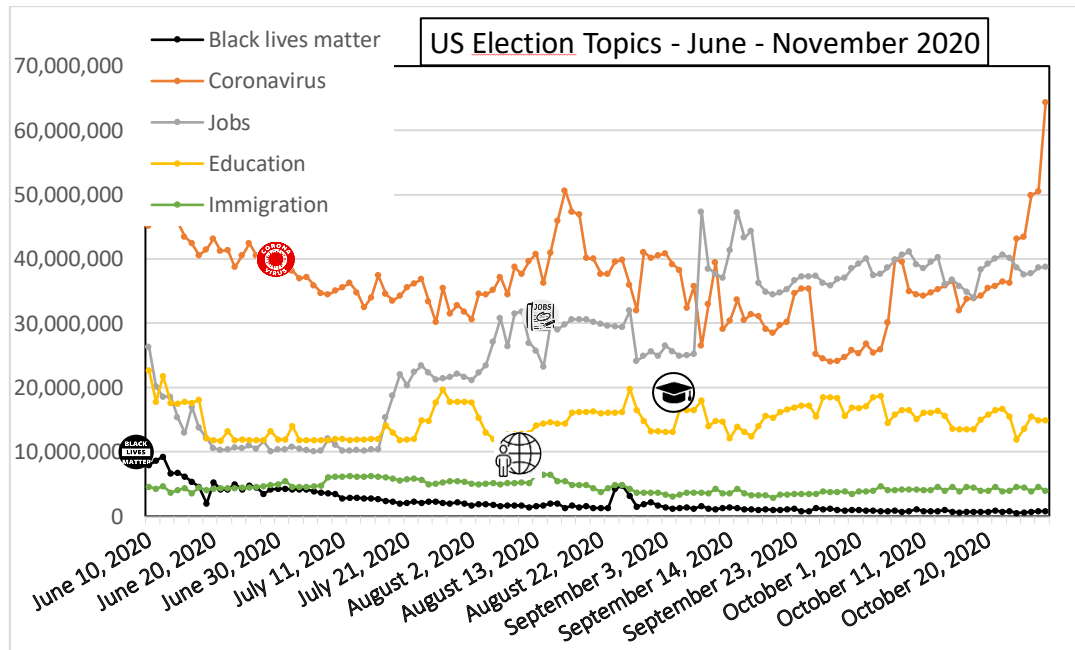

Figure 3. Evolution of some topics during the 2020 US electoral campaign (June–October)

Usually, the crowd's interest in a given topic is relatively stable during the electoral campaign as shown with the *Education* or *Immigration* themes. For others, the demand pattern can vary as, for instance, the topic *Coronavirus* corresponding to a decrease up to the end of July, a spike in the beginning of August followed by a decline up to the first week of October and finishing with a significant rise at the end of October. This line is correlated with the number of corona deaths per day, being 763 in June, and rising to 974 in August. A reduction was observed in September to 778 deaths before a rebound in November (1,244 deaths/day).

In Figure 3, the unemployment or *Jobs* curve presents an increase during this time period, starting around 10M queries in mid-June to 40M by the end of October. One must recall that the unemployment rate reached 14.7% in April 2020 compared to 3.6% in January. Of course, the Trump administration decided to act against this issue. They injected the first federal stimulus in March 2020 (a check of $1,200 for each eligible adult). A second federal subsidy was created with an extra $600 per-week bonus for jobless persons, but this support expired at the end of July without being prolonged. Unsurprisingly, the US citizens were then really concerned with this joblessness crisis, even though the unemployment rate decreased to reach the level of 7.9% in September and 6.9% in October.

For a relatively shorter timespan (a few weeks), US online users demonstrated an interest in specific flame topics such as *Black Lives Matter*. As shown in Figure 3, this theme was searched intensively during a few weeks, then returned to a low level of attention. An explanation of this phenomenon is related to G. Floyd's death (May 25th) in the beginning of the graph and then to J. Blake's shooting (August 23rd).
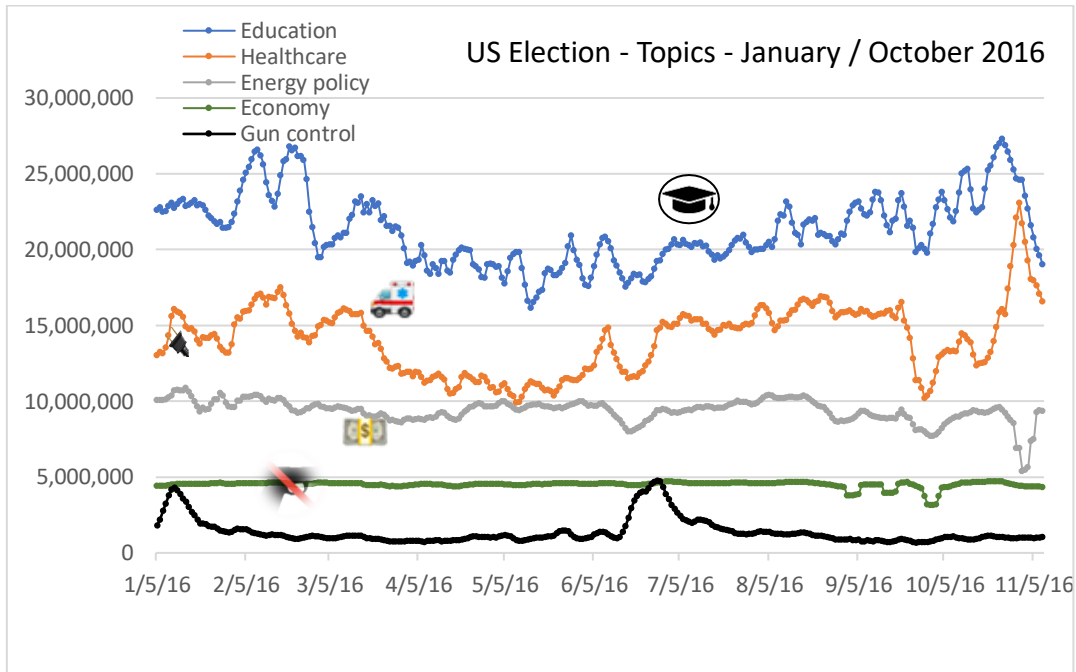

Figure 4. Evolution of the four topics during the 2016 US campaign

When analyzing the main topics of the 2016 US election, the most popular topics were *Education*, and *Healthcare*, while *Energy Policy* and *Economy* received less attention from the US internet users. The underlying demand patterns can however vary as, for instance, the topic *Economy* corresponding close to a horizontal line while *Healthcare* exhibits some fluctuations reflecting its volatility as a topic. In the bottom of Figure 4, the *Gun control* theme also fluctuates as a flame topic. The spike can be explained by a terrorist attack (Orlando, FL, June 12th).

In 2016, the topic *Immigration* received relatively little attention from the US Internet users. To explain this, we can put forward the argument that this topic is discussed more intensively in social networks and does not completely correspond to a classical Internet search.

As for the previous analysis, one can verify whether the monthly differences between the topical curves are significant or otherwise. To achieve this, the *t*-test [37] with a significance level of 1% has been applied for all topics depicted in Figures 3 and 4. These differences are always significant except for a few cases. In Figure 3, between the topics *Coronavirus & Jobs*, the differences cannot be deemed significant. Secondly, the level achieved by the subjects *Gun Control & Economy* in Figure 4 must be analyzed as similar for the months of January through March as well as for June through October.

5. CANDIDATES' TWEETS ANALYSIS

During an election, the candidates aim at conveying their leadership capabilities to the public by adopting an appropriate style and rhetoric. Moreover, they need to discuss and present solutions

to concerns judged important by the voters. To analyze these differences between both candidates, tweets sent during the last two months of the 2020 campaign have been automatically analyzed using stylometric methods (Savoy, 2020).

This corpus is composed of 2,278 tweets sent by Trump together with 871 written by Joe Biden. In addition, we have considered the 924 tweets sent by Trump during the last two months of the 2016 election. One must underline that due to the pandemic situation, the social networks, and particularly Twitter, could be considered as the first channel of communication in 2020. Moreover, the subjects discussed during an interview or TV debates are not fully controlled by the candidates. Therefore, the content of social networks represents more precisely the topics judged essential by each nominee.

Moreover, using Twitter, both candidates can reach their supporters directly and instantaneously. In this view, Trump sent, in mean, 37 tweets per day to around 89 million followers, achieving a higher level than Biden with 14 tweets per day broadcasted to 19 million supporters. Of course, not all followers are US citizens who will vote, but the Republican nominee shows a larger potential. In 2016, Trump only sent 15 tweets per day, indicating the increasing importance of Twitter in the 2020 election.

Analyzing tweets' content, it is a surprise to observe that 51.7% of Trump's tweets are retweets (compared to 3.5% in 2016). In addition, many tweets are repeated many times as, for example, "Law and order" occurring 14 times in Trump's thread. To emphasize the importance of some words, the Republican nominee wrote 5.4% of them in uppercase (e.g., AMERICA, GREAT, VOTE) (7.2% in 2016). Trump opts for shorter words than Biden; Only 16.7% of Trump's words count more than six letters compared to 19.8% for Biden. According to Lakoff & Wehling [38], short words are easier to understand and remembered. Selecting short words when possible was also recognized by Sorensen, the main ghostwriter of President Kennedy, who wrote:

"Lincoln never used a two- or three-syllable word where a one-syllable word would do" [39].

Including emotional words is an important rhetoric characteristic because by making people anxious, they will pay careful attention to the message [40], [41]. In addition, emotions could even detach people from their party affiliation and their constrictive ways of thinking [42].

Such a choice usually generates negative campaigning and can explain Trump's success in the 2016 election [43], [20]. At that time, the four main targets were the immigrants viewed as criminals, China destroying jobs in America, the political establishment in Washington D.C. ignoring the grassroots people and the media reporting fake news to support the country's elite. In 2020, Trump only considers the last target, and particularly CNN (appearing in 3% of the presidential tweets). References to China occur in only 2.5% of his tweets, while immigrants are mentioned in only 1%. To provoke fear and angering, Trump selected a new target in 2020. He plays up the risk of large riots, lootings and uprisings if the radical left win in this election (9.8% of his tweets). To support his view, Trump highlights the protests generated by G. Floyd's death (May 25th) and J. Blake's shooting (August 23th).

In response, the Democratic nominee did not opt for an intensive negative campaign but chose to put forward the prominence of moral values (e.g., honesty, love) and symbolic notions (e.g., truth, faith, justice). Biden then compared these values with those assumed to represent Trump's intents and actions (e.g., fear, hate, violence). In this perspective, symbolic terms appear in 39% of Biden's tweets and moral values in 21%. To more directly involve the reader, Biden's tweets contain several denominations associated with people (e.g., Latinos, folk, Americans, women) (30% of his tweets).

The large interest shown in the corona virus in Figure 3 is not fully taken into account by the two candidates. Of course, this situation is problematic for Trump's administration, judged as the "greatest failure" by Biden. Terms related to covid-19 only appear in 3.4% of Trump's tweets but in 12.1% of those written by the Democratic nominee. For the jobs and unemployment topic, the occurrence rate is 5.5% of tweets sent by Trump and 6.5% by Biden. These percentages could be viewed as low, but one must recall that many tweets are sent to indicate a rally of the candidate in a given city or to support a candidate in a race to the Senate or the House of Representatives (e.g., "Kat has my Complete and Total Endorsement!")

However, citizens are concerned with other topics as shown in Figure 3. In tweets sent by both candidates in September and October 2020, one can count only six occurrences of words beginning with "immigr*" or "education". The word "energy" was mainly used to qualify Biden (e.g., "low energy Joe" in tweets sent by Trump) and not to discuss the energy issue. As other examples, one can cite "climate" appearing 27 times in tweets written by Biden, and never in those sent by Trump. During an electoral campaign, candidates do not speak about all pertinent political subjects and the terms "budget," "debt" or "deficit" occur rarely. This 2020 election is not an exception and the word "budget" never appears. The word "debt" (three occurrences) is related to the issue of student debt or to the expression "debt of gratitude". When using "deficit" (two occurrences), Trump tweets about "reduce deficit through cuts to social security" while Biden talks about the trade deficit with China.

To illustrate the rhetoric differences between the two candidates of the 2020 election, we have selected two of the most specific tweets of each candidate (the applied method is described in [5]). With Trump, the following two examples provide a clear picture of the presidential style during this campaign with lies and expressions written to generate fear and anger.

> "RT @realDonaldTrump: Wow!!! NOW DO SOMETHING ABOUT THIS, THE BIGGEST OF ALL POLITICAL SCANDALS (IN HISTORY)!!! BIDEN, OBAMA AND CROOKED HILLARY LED THIS TREASONOUS PLOT!!! BIDEN SHOULDN'T BE ALLOWED TO RUN - GOT CAUGHT!!!" Trump, October 8th, 2020.

> "RT @DonaldJTrumpJr: "Peaceful Riots" "Peaceful Looting" and "Peaceful Arson" Trump, September 7th, 2020.

With Joe Biden, the two representative tweets focused on the pandemic situation, denoting the breakdown of the current administration. In the second tweet, the democratic nominee wrote with positive terms in an assertive tone (we will solve this issue).

> "Donald Trump is holding a rally in Michigan today—while also refusing to fully fund their National Guard. These are frontline workers against COVID-19, responsible for testing and distributing food and medical supplies. Michiganders need a pandemic response—not a pep rally. https://t.co/UkEqVv9zXe"Biden, September 10th, 2020.

> "We have to heal this nation. https://t.co/wD45ZGcKO2" Biden, October 16th, 2020.

6. CONCLUSION

First, this monitoring study of the last US presidential election reveals, in real time, the popularity of the two candidates based on the number of search queries sent to numerous search engines and social networks. As shown in Figure 1, Donald Trump received a larger number of requests than

the Democratic nominee Joe Biden. The difference between them is significant even when considering that Trump, as the incumbent president, is the object of some additional enquiries over a normal candidate.

Moreover, when compared to the 2016 US election (see Figure 2), Trump in 2020 achieved a very high level, particularly after October 1st 2020 when he sent a tweet announcing he caught the covid virus. When comparing the two Democratic candidates in the last two elections, one can see that Hillary Clinton was able to obtain more queries than Joe Biden. This fact could justify that the 2020 US election was more against Trump than in favor of Biden.

Second, through a selection of topics, we could observe the evolution of the people's interests during the 2020 (Figure 3) or 2016 elections (Figure 4). Clearly, in 2020 the US Internet users demonstrated a noteworthy interest in both coronavirus and unemployment issues. This can be explained by knowing that, from June to November 2020, this virus killed around 890 people per day, and that the employment rate was around 8.5%. Figure 3 indicates that other subjects have captured the attention of citizens such as education or healthcare, both being clearly important during the 2016 election (see also [44]).

For other topics, an event could increase the interest for a few days or weeks before returning to a low level. This was the case for "gun control" in 2016 or "Black lives matter" in 2020. Even if the newspapers or TV shows are talking to these issues, they cannot be considered as essential in a long-term campaign.

Third, due to the pandemic situation in 2020, Twitter had played a significant role in motivating the supporters and convincing the independents. In this perspective, Trump's rhetoric emphasized the violent riots in several US cities during the summer, a situation generating fear and anguish in the people. Trump is then perceived as the best safeguard against those uprisings and insurrections. In addition, the Republican nominee insisted on his capabilities to reduce the unemployment rate and to create new jobs. On the other hand, the Democratic candidate wrote tweets underlining the prominence of moral values (e.g., hope, honesty) and symbolic notions (e.g., democracy, justice). Biden put forward the incapacity of Trump's administration to solve both the pandemic and unemployment. He was also the only one to talk about the climate change issue.

During the 2020 campaign, both Trump and Biden opted to frequently tweet on two main problems (coronavirus and jobs) and to support their candidates to the Senate and House races. Thus, they are not talking on other important topics such as education, healthcare insurance, education, or immigration. Foreign policy, trade, or family are other subjects not being discussed by the two main candidates, leading to an electoral campaign limited to a few topics.

ACKNOWLEDGEMENTS

REFERENCES

[1] Graham, S., Milligan, I., & Weingart, S. 2016. Exploring Big Historical Data. The Historian's Macroscope. Imperial College Press, London (UK).
[2] Brin, S, & Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. Computer Network and ISDN Systems, 30, 107-117.

[3]   Labbé, D., & Monière, D. 2003. Le discours gouvernemental. Canada, Québec, France (1945-2000). Honoré Champion: Paris.

[4]   Labbé, D., & Monière, D. 2008. Les mots qui nous gouvernent. Le discours des premiers ministres québécois: 1960-2005. Monière-Wollank: Montreal.

[5]   Labbé, D., & Savoy, J. 2021. Stylistic analysis of the French presidential speeches: Is Macron really different? Digital Scholarship in the Humanities, to appear.

[6]   Pauli, F., & Tuzzi, A. 2009. The end of year addresses of the presidents of the Italian republic (1948–2006): Discourse Similarities and Differences. Glottometrics, 18, 40–51.

[7]   Savoy, J. 2015. Vocabulary growth study: An example with the State of the Union Addresses. Journal of Quantitative Linguistics, 22, 289-310.

[8]   Savoy, J. 2015. Text clustering: An application with the State of the Union addresses. Journal of the American Society for Information Science and Technology, 66, 1645-1654.

[9]   Kubát, M., & Cech, R. 2016. Quantitative analysis of US presidential inaugural addresses. Glottometrics, 34, 14-27.

[10]   Sylwester, K., & Purver, M. 2015. Twitter language use to reflects psychological differences between Democrats and Republicans. PLoS One, 10(9).

[11]   Raubach, E.E. 2019. Does political ideology influence how politicians speak? A linguistic analysis of one-minute speeches from the US House of Representatives of the 115th Congress. Document numérique, 22(1-2), 127–138.

[12]   Yu, B. 2008. Classifying party affiliation from political speech. Journal of Information Technology and Politics, 5, 33-48.

[13]   Yu, B. 2013. Language and gender in congressional speech. Literary and Linguistic Computing, 29,118-132.

[14]   Pennebaker, J. 2011. The secret life of pronouns. What our words say about us. Bloomsbury Press: New York (NY).

[15]   O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. 2010. From tweets to polls: Linking text sentiment to public opinion time series. Proceedings 4th International AAAI Conference on Weblogs and Social Media, 122-129.

[16]   Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi Y., Cardie, C., Riloff, E., & Patwardhan, S. 2005. OpinionFinder: A system for subjectivity snalysis. Proceedings Empirical Methods for Natural Language Processing, Vancouver (BC), 34-35.

[17]   Mohammad, S.M., Zhu, X., Kiritchenko, S., & Martin, J. 2015. Sentiment, emotion, purpose, and style in electoral tweets. Information Processing & Management, 51, 480-499.

[18]   Arnold, E., & Labbé, D. 2015. Vote for me. Don't vote for the other one. Journal of World Languages, 2, 32-49.

[19]   Young, L., & Soroka, S. 2012. Affective news: The automated coding of sentiment in political texts. Political Communication, 29, 205-231.

[20]   Hart, R. 1984. Verbal style and the presidency. A computer-based analysis. Academic Press: Orlando (FL).

[21]   Bligh M., Merolla J., Schroedel J., & Gonzalez, R. 2010. Finding her voice: Hillary Clinton rhetoric in the 2008 presidential campaign. Woman's Studies, 39, 823-850.

[22]   Hart, R. 2020. Trump and us: What he Says and Why People Listen. Cambridge University Press: Cambridge (UK).

[23]   Tausczik, Y., & Pennebaker, J. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29, 24-54.

[24]   Slatcher, R., Chung, C, Pennebaker, J., & Stone, L. 2007. Winning words: Individual differences in linguistic style among U.S. presidential and vice presidential candidates. Journal of Research in Personality, 41, 63-75.

[25]   Lichtman, A. 2016. Predicting 2016: How the 13 keys to the White House will turn in 2016. HubPages, June 7th. URL: hubpages.com/politics/Predicting-the-2016-Election-the-13-Keys-to-the-White-House.

[26]   Granka, L. 2013. Using online search traffic to predict US presidential elections. Political Science and Politics, 46, 271–279.

[27]   Hargittai, E. 2015. Is bigger always better? Potential biases of big data derived from social network sites. Annals of the American Academy of Political and Social Science, 659, 63-76.

[28] Boller, P. Jr. 2004. Presidential campaigns. From George Washington to George W. Bush. Oxford University Press: Oxford.

[29] Eisenstein, F. J. 2019. Introduction to Natural Language Processing. The MIT Press: Cambridge (MA).

[30] Glauser, C., Schmid, L., Kurpicz-Briki, M. 2020. 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 (ICEGOV2020). The ACM Press: New York.

[31] Cleverdon, C., Mills, J, & Keen, M. 1966. Factors determining the performance of indexing systems. Technical Report, Cranfield (UK), ASLIB Cranfield Research Project.

[32] Ehrenfreud, M., & Tankersley, J. 2016. Seven things to know about Hillary Clinton and Donald Trump before you vote. Washington Post, November 7th.

[33] Furnas, G., Landauer, T., Gomez, L., & Dumais, S. 1987. The vocabulary problem in human-system communication. Communications of the ACM, 30, 964-971.

[34] Manning, C., Raghavan, P., & Schütze, H. 2008. Introduction to Information Retrieval. Cambridge University Press: Cambridge (MA).

[35] Hart, R., Childers, J, & Lind, C. 2013. Political Tone. How Leaders Talk and Why. The University of Chicago Press: Chicago (IL).

[36] Hart, R.P., & Curry, A.L. 2016. The third voice of America politics. Presidential Studies, 46(1), 73-97.

[37] Grimm, L.G. 1993. Statistical Applications for the Behavioral Sciences. John Wiley & Sons: New York.

[38] Lakoff, G., & Wehling, E. (2012). The Little Blue Book: The Essential Guide to Thinking and Talking Democratic. Free Press, New York.

[39] Carpenter, R.H., & Seltzer, R.V. (1970). On Nixon's Kennedy Style. Speaker and Gavel, 7, 41-43.

[40] Marcus, G., MacKuen M.B. 1993. Anxiety, enthusiasm, and the vote: The emotional underpinning of learning and involvement during presidential campaigns. American Political Science Review, 87(3), 672-685.

[41] Ansolabehere S, & Iyengar S. 1995. Going Negative: How Attack Ads Shrink and Polarize the Electorate. New York: Free Press.

[42] Brader, T. 2005. Strinking a responsive chord: How political ads motivate and persuade voters by appealing to emotions. American Journal of Political Science 49(2), 388-405.

[43] Sides, J., Tesler, M., & Vavreck, L. 2018. Identity crisis: The 2016 presidential campaign and the battle for the meaning of America. Princeton University Press: Princeton.

[44] Glauser C., Savoy J., Schmid L. 2021. User searches on multiple channels during the US presidential election 2020. Journal of Political Sciences & Public Affairs, 9(3), Article #3.

APPENDIX

The user data from more than 1,400 selected search engines and social media networks can be collected through API interfaces. These APIs are programmed in individual standard Java applications collecting publicly available data provided by the suppliers themselves or, for some sources, with privileged access to the index data (in few cases, such an access is subject to a fee). The user demand data were collected and then validated (different parameters are considered here, e.g., time-span, country extensions, user references, etc.).

To illustrate this aspect, Tables A.1 ("Donald Trump") and A.2 ("Joe Biden") report the frequency values obtained with these two queries and seven selected sources.

Table A.1. Query "Donald Trump" with US sources (sent Oct. 5th, 2020).

Source	Yahoo!	Reuters	Google	Facebook	Twitter	Bloomberg
Frequency	1,027,929	466,636	2,917,697	908,083	781,950	1,220,383

Table A.2. Query "Joe Biden" with US sources (sent Oct. 5th, 2020).

Source	Yahoo!	Reuters	Google	Facebook	Twitter	Bloomberg
Frequency	46,210	169,152	287,820	195,525	115,708	253,175

AUTHORS

Dr. Christoph Glauser studied mass media, political science and history at the universities of Berne and Geneva (1985-1992). In 1994 he found his own "Institute for applied argumentation research IFAAR" in Berne, where he started doing scientific research about national and international election campaigns and content analysis for media and news agencies. From 1996 to 2000 he directed a research program at the Swiss Federal Institute of Technology (ETH) in Zurich about "expert communication on biotechnology in the public sphere and in the media". Between 1998 and 2007 he was lecturing as a guest Professor at Communication School of the University of Washington Seattle USA. Since 2001 Dr. Glauser has been director of the "IFAAR" institute and managing director of ArgYou AG. He has developed the "find-engine" for measuring "user audience search volumes" to show the impact of online campaigns, PR and e-government.

Prof. Jacques Savoy is full Professor in Computer Science at University of Neuchatel (Switzerland). After received a Ph.D. in quantitative economics from the University of Fribourg in 1987. From 1987-92 he was member of the faculty of Computer Science at the University of Montreal (Canada). His research interests cover mainly natural language processing and particularly information retrieval for other languages than English (European, Asian, and Indian). He participated from many years to various evaluations campaigns (TREC, CLEF, NTCIR, FIRE) dealing with these questions. His current research interests are related to statistical modeling and evaluation of natural language processing such as text clustering and categorization as well as authorship attribution. As application domain, he is working on political speeches (both governmental or electoral) with a particular focus on the United States.

Loris Schmid was born in 1992 in Visp, Switzerland. He studied g at the University of Berne where he attained a Master of Science in Economics. After an internship at the IFAAR institute, he now works as a Research Assistant at IFAAR. During the UMUSE (User Monitoring of the US Election) project, Loris Schmid was employed by the University of Neuchâtel from August 2020 until February 2021 performing data analysis and processing.