

Sponsored Search for Political Campaigning during the 2008 US Elections

Eni Mustafaraj
Department of Computer Science
Wellesley College
Wellesley, MA
emustafa@wellesley.edu

Panagiotis Takis Metaxas
Department of Computer Science
Wellesley College
Wellesley, MA
pmetaxas@wellesley.edu

ABSTRACT

We have collected a set of 1131 textual ads that appeared in the Google Search results when searching for a candidate name running in the 2008 US Congressional elections. We have categorized the advertisers in four different categories: commercial, partisan, non-affiliated, and media. By analyzing the content of the collected ads, we discovered that the majority of them (63%) are commercial ads that have no political message, while the partisan group contributed only 14% of the ads. Furthermore, only 21 out of 124 monitored candidates were actively participating in sponsored search, by providing their own political message. We describe the different ways in which the advertisements were used and several problems that damage the quality of sponsored search, providing some suggestions to avoid such issues in the future.

Keywords

sponsored search, content analysis, 2008 US elections

1. INTRODUCTION

The last report “The Internet and the 2008 Election” [1], published by Pew Internet & American Life Project, found out that 39% of Americans have used Internet to access “un-filtered” campaign materials during the 2008 primary elections. Since the search for information on the Web usually begins with queries in a search engine, the results produced by the search engine could have an impact on the kind of opinion an individual might form about a candidate. In the framework of a project aimed at capturing efforts of manipulating search engine results for political reasons, during a 6 months period (June – December 2008), we collected once a week search results targeting the names of more than a hundred candidates for the US 2008 Congressional elections, who were reportedly involved in crucial and highly contested races. Our analysis of the organic search results has been reported in [2], thus, in this paper we will focus on the sponsored search results only.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IRA '09 July 23, 2009. Boston, MA, USA

Copyright 2009 ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

Table 1: Distribution of ads during the 11-days period of data collection.

Date	Time (EDT)	# Ads	# Cand.
2008-10-27	09:00 - 12:00	73	47
2008-10-28	15:00 - 18:00	65	48
2008-10-31	08:30 - 11:30	119	61
2008-11-04	11:00 - 14:00	119	71
2008-11-05	12:00 - 15:00	68	43
2008-11-06	08:15 - 11:15	73	50
2008-11-07	12:00 - 15:00	72	49
2008-11-11	12:00 - 15:00	80	57
2008-11-14	08:15 - 11:15	95	56
2008-11-24	11:15 - 14:15	176	84
2008-12-01	08:20 - 11:20	191	90

According to [3], the Google search engine attracts more than 70% of the query volume¹ in the United States, therefore, we have limited our data collection process to the results returned by Google. When using the Google API (Application Programming Interface) to automatically access the Google index in order to get results for a given query, the API will return the organic search results only. This is different from issuing a query to the Google web interface, because the latter will return an HTML page that usually contains several other types of results: advertisements at the top of the organic results, advertisements at the side of results, and interspersed in the page, results from searching news, videos, blogs, books, shopping, etc., or occasionally a list of related query search phrases in the bottom². Since Google suggests the use of its API for automatic collection of search results, our weekly experiments contain only the organic search results. However, during the period Oct. 27 - Dec. 01 2008, in 11 different days, we automatically collected search results from the web interface, taking care to distribute the queries during a 3-hour period. The number of ads collected during these days, as well as the number of candidate names targeted by these ads are summarized in Table 1.

Reportedly [5], Barack Obama spent one million dollars on Google AdWords, in February 2008. Moreover, the indica-

¹Although other sources report different numbers, all sources concur that Google is by a large margin the market leader in search.

²On May 12, 2009, Google announced another change to its Search service, called Search Options [4]. Therefore, the above description might not apply anymore to the appearance of search results in a web page.

tions for the gubernatorial races taking place in Fall 2009 are that the campaigns and issue groups are making an earlier and more sophisticated use of AdWords than the presidential candidates at the comparable stage [6]. Thus, AdWords is already becoming an important tool for online political campaigning, with the potential to surpass other types of advertisements in the future, because users are also moving en-masse toward online political information gathering. From this viewpoint, we think it is of value to look retrospectively on how sponsored search performed during the 2008 elections, in order to uncover potential issues that need to be carefully examined in the future. To our best knowledge, such an exploratory analysis has not appeared previously in the research literature. We recognize that our methodology for data collection has flaws, therefore, we suggest how to improve this process in the section of future work.

2. THE DATA COLLECTION PROCESS

During a careful selection process, described in [2], we had selected 60 congressional races that were predicted to be highly contested. Later on, for the purpose of collecting advertisement data, we added the two pairs of candidates in the presidential race. Thus, there were 124 candidates (120 candidates for the Congress and the 4 candidates for president/vice president) followed by our experiment. Each candidate name (in quotation marks) was issued as a query to the Google Search Web Interface, and the returned HTML pages containing organic and sponsored search results were stored. By parsing the HTML code, it was possible to detect and extract sponsored search results, when present. In total, we collected 1352 HTML files (during the first 3 days of collection, the 4 presidential candidates were not included, $3 * 120 + 8 * 124 = 1352$), and by parsing these files we extracted 1131 advertisements. By aggregating all extracted data, it resulted that 112 of the 124 candidates did have at least one advertisement targeting their name.

Differently from advertisements within websites, which are usually visually stimulating, sponsored search ads are purely in textual form. Commonly, a textual ad is a triplet containing a caption, a short message, and a URL. When using Google Search, textual ads are displayed either in the top of the page (right above the top ten list of results) or in the side. Google Search displays up to 3 ads in the top section and up to 8 ads in the side section. We counted 72 ads in the top position compared to 1059 ads in the side position. This fact shows that the quality score used by Google to determine an ad position has ranked as non-top quality the majority of the ads. In the following section, we discuss several findings of our exploratory content analysis.

3. FINDINGS

By extracting the captions and URL portion of each textual ad, we found out that in the collected set of **1131** ads, there are **489** unique ads (i.e., their caption is unique) contributed by **147** websites (advertisers). After analyzing the content of the ads and the related URLs (we visited several websites, whose nature was not obvious from the ad text), we grouped the websites in four categories: **commercial** (websites that sell products or information), **partisan** (websites promoting or opposing candidates and their agendas), **non-affiliated** (websites contributing political and electoral information, but not affiliated with the candidates or their

parties), and **media** (websites of newspapers, TV stations, magazines, etc.). Statistics about these four groups are summarized in Table 2. As the numbers indicate, 63% of the ads are purely commercial and have no political meaning. The following subsections discuss the content of ads for each category.

3.1 The Partisan Category

The analysis of the partisan category showed that only 21 candidates (17 running for the congress and the 4 for the presidency) had actively included ads directing to their campaign websites, with text messages such as:

Support Senator John Sununu. Join Team Sununu Today!
 Darcy Burner is exactly what we need in Congress!
 Join John McCain's Team and Contribute Today! Thank You
 Tom Udall needs your help to keep fighting for New Mexico.

Besides these direct, positive ads, there are two other types of ads.

3.1.1 Targeting the Opponent Name

Some candidates had bought ads that targeted the name of their opponent. For example, when the name of Gerry Connolly was searched for, an ad from his opponent appeared, with the caption "Keith Fimian for Congress". One other example was that of the democratic candidate Tom Udall, whose ad "Tom Udall for New Mexico" appeared when searching for the republican candidate who was running for the house seat in his district, Darren White, although Tom Udall himself was racing for the senate seat. While these examples can be labeled as "awareness ads" (giving to people searching for a candidate the possibility to know about the opponent), there were occasions of negative ads targeting the opponents, although not directly from the candidates. For example, when searching for the candidate Norm Coleman, several negative ads directed to his opponent Al Franken would appear, such as "Franken: Unfit for office" or "Is Al Franken lying?", contributed by a website named www.FranklyFranken.com.

3.1.2 Negative Ads

The more frequent use of negative ads was to directly target the name of a candidate. We found 25 negative ads that have appeared 67 times, targeting 16 candidates. As negative ads we treated those ads that contained disapproving language toward the candidate. Some examples of negative ads are shown in Table 3. However, only 4 of these ads (for a total count of 11 appearances) were published by the candidates' websites, all the other negative ads came from other websites, not directly affiliated with the candidates. The most worrisome fact about the negative ads is that they originate from websites which were shut down directly after the elections, and that have used the candidates' names in their URLs. Some of those now extinct websites are:

ShaheenForSenate.com, VoteMcNerneyOut.com,
GilibrandUnfiltered.com, TheRealBobRoggio.com.

Their sudden disappearance might be an indicator of unfair and unethical political campaigning.

3.2 The Non-affiliated and Media Categories

Non-affiliated sites use the names of the candidates to attract traffic to their sites, by using template ads such as:

Table 2: The four categories of websites, which ads were displayed in response to candidate names.

Website type	# sites	# targ. cand.	# ads	%
<i>Commercial</i> (e.g., amazon.com , public-records-now.com)	79	92	712	63%
<i>Partisan</i> (e.g., BarackObama.com , dccc.org)	37	39	157	14%
<i>Non-affiliated</i> (e.g., theMiddleClass.com , houseRaceTracker.com)	16	72	206	18%
<i>Media</i> (e.g., newser.com , FoxNex.com)	15	20	56	5%
Total	147	112	1131	100%

Table 3: Examples of negative ads. Ad text, including errors, appears verbatim.

Ad Caption	Ad Text
Who’s the real Bob Roggio	Toxic Chemicals. 500 PA Jobs Lost. Now Wants \$2,031 in Higher Taxes!
McNerney Undermines War	A Sellout Democrat for Al Qaeda Opposed Reinforcements in Iraq
Women Against Sarah Palin	Anti-choice, anti-gay, pro-drilling Palin does not speak for women.
Lobbyist Steve Stivers	Career Lobbyist Steve Stivers for Congress

View Rep. [Candidate Name]’s middle-class voting record here.
 Do you agree with [Candidate Name] on the issues that matter to you?

For example, the website [themiddleclass.com](#) alone targeted 45 candidate names for a total of 110 ad occurrences, with the majority of them (around 75%) past the election date. The non-affiliated sites together targeted 72 candidate names for a total of 206 ad occurrences.

Media websites also use the same strategy, by using the candidate names in the caption of the ad, for example:

[Candidate Name] Fox-NEWS *provided by foxnews.com*
 [Candidate Name] News *provided by examiner.com, news.aol.com, etc.*

The media category was the smallest one in the dataset. It had 15 different advertisers, targeting 20 candidate names for a total of 56 ad occurrences.

3.3 The Commercial Category

The commercial ads outweigh by far all other types of ads. In fact, we counted 712 occurrences of ads targeting 92 candidates. The first reason for the abundance of the commercial ads is that three commercial websites, specialized in finding people:

[public-records-now.com](#)
[usa-people-search.com](#)
[wink.com](#)

have contributed together 380 ad occurrences, targeting 71 candidate names. The second reason is that several candidates have names shared by other individuals that have monetizable professions. That happened to the candidate Brian Davis that had an artist doppelganger attracting 59 ads from online sellers of art, to the candidate Dan Seals that had a country singer doppelganger, or to Steve Greenberg that had a record producer doppelganger. Furthermore, some of the candidates themselves have in their course of life produced books or music, played professional sports, etc., so that websites such as [Amazon.com](#) and [eBay.com](#) will target their names. And yet another source of ads were sites trying to cash in with products not authorized by the candidates, as the examples in Table 4 show.

Very often, an ad will not target any of the candidates in a race. In fact, 25% of all ad occurrences were such mistargeted, commercial ads, because they used various spellings

of the candidate names. The most common use of such ads were from the people finder websites, which will target any variation of a name and surname, or even have ads targeting only the first name, such as: “Find Dean”. While this kind of ads at least understands that the user is searching for a person name, there are other ads that miss the semantic category of the query. For example, when searching for the candidate Baron Hill, appeared the ad: “Barton Hills home 4 Sale”, where Barton Hills is a geographic location; when searching for Darren White, appeared the ad: “White Pages directory”, when searching for Victoria Wulsin appeared ads from the Victoria’s Secret company, when searching for Nancy Boyda appeared the ad “Local Nannies”, and when searching for Mark Begich appeared several ads related to companies or products with the keyword “mark”, such as “Vanmark Collectibles”, “TriMark Corporation”, “Primemark”, or “Colormark at Amazon.com”. An interesting fact is that although there are five candidates with their first name Mark, only one of them, Mark Begich, attracted the mentioned ads.

4. DISCUSSION AND SUGGESTIONS

As in traditional media, the ads appearing in the sponsored search results were used to both promote or oppose a candidate’s political message. However, there is a big difference. Advertisements in TV, radio, and newspapers clearly indicate who has paid for it, by mentioning this fact in the ad, so that voters know who is responsible for the message. Political online advertisements has yet to be regulated by law, however, the Federal Electoral Committee (FEC) in 2006 has advised [7] the following about Internet advertisements:

Because Internet advertisements are public communications, an individual or group must include a disclaimer on any Internet advertisement that expressly advocates the election or defeat of a clearly identified Federal candidate, or on any Internet advertisement that solicits contributions.

None of the ads we collected and analyzed had any disclaimer in their text. Often several clicks in the landing page are needed to find out who is responsible for the advertised

Table 4: Examples of commercial ads.

Ad Caption	Ad Text
Obama 4-Coin Collection	Tribute to 44th President of US in velvet gift box. Perfect gift!
President Obama T-Shirts	Yes We Did! Welcome Mr. President. Tees, Stickers, Buttons, Yard Signs
Barack Obama	Shop our Best Designs or Create Your Own Barack Obama Merchandise!
Palin Humor Tees	Get Sarah Palin Parody Shirts, Stickers, Buttons, & Gear

message, and in the occasion of extinct websites, this is not possible at all. While it is true that for such short ads is difficult to have an extra disclaimer message, we believe that it might be possible to circumvent this drawback, for example, by including a small icon that indicates a pro or contra message.

Additionally, there are several other improvements that search engines can undertake in order to make sponsored search results more effective and help voters to avoid confusion or ambiguity. Given the fact that elections take place in two-year intervals and candidate names are known several months before the election date, there is plenty of time for search engines to put such measures into actions.

1. Maintain a list of candidate names registered with FEC. For searches matching names in this list, avoid displaying ads that do not contain the correct spelling of the candidate name either in the caption or in the text message. In this way, all spurious ads resulting from broad matching are eliminated.
2. Require a disclaimer for all ads targeting the names of political candidates.
3. Resolve cases of several individuals with the same name (similar to the Wikipedia disambiguation page for people's names).
4. Clearly mark political ads (for example by using a background color) to distinguish them from commercial or informational ads.

We think that such measures are in the interest of search engines. By taking actions to ensure that only relevant ads are delivered, more candidates, political groups, or others might be interested to advertise. By providing recognition cues and categorizing ads properly, the users might be encouraged to click more often on such ads.

5. FUTURE WORK

Our data collection methodology has some flaws. We searched for less than one fourth of the candidates (though that subgroup was involved in highly contended races); we collected data only for a few days before elections, only once a day, and for one geo location only (that of our server). Furthermore, our queries contained the names of the candidates only, whereas the candidates might have been advertising with many other words, such as "Illinois candidates for senate", "elections in Montana", "economic crisis", "war in Iraq", etc. Coming up with a comprehensive set of keywords in advance is something that requires much more preparatory work, might need the involvement of a political analyst, and also a constant monitoring of the regional query trends volume using the Google AdWords tools.

Another important area of future study is that of contextual advertisements, which appear on third-party websites via, for example, the Google AdSense technology. In our corpus of 65,700 HTML pages, collected by storing the top 20 search results returned by Google when searching with the candidate's names over a 6 month period, 20% of the pages have in their HTML code the *google_ad_client* script that delivers the ads. Because the ads are delivered in real time and are not part of the HTML source of the page, we were not able to analyze their content. We need thus means to automatically record ads delivered within a webpage at a specific point in time, in order to analyze their content. Even for the candidates themselves it is important to know on what websites their advertisements are being displayed, because they might not agree to become affiliated with websites that use derogatory language. In particular, John McCain's campaign did withdraw several of his ads from some anti-Obama websites, when it was discovered that such websites were using offensive language toward Obama [8].

All these issues need to be addressed carefully, if a more comprehensive analysis of online advertisement for political campaigning is desired.

6. REFERENCES

- [1] Online Report, *The Internet and the 2008 Election*, <http://www.pewinternet.org/Reports/2008/The-Internet-and-the-2008-Election.aspx>
- [2] P. Metaxas and E. Mustafaraj, *The battle for the 2008 US Congressional Elections on the Web*. Proc. of WebSci-09: Society On-Line, March 18-20, 2009, Athens, Greece.
- [3] StatCounter - Global Stats. Last access on July 10, 2009. http://gs.statcounter.com/#search_engine-US-daily-20080701-20090610
- [4] The Official Google Blog, *More search options and other updates from our Searchology event*, May 12, 2009. <http://googleblog.blogspot.com/2009/05/more-search-options-and-other-updates.html>.
- [5] S. L. Stirland, *Our Brand Is Crisis: Prez Candidates Buy Words To Brand Each Other Online*, Sep. 19, 2008. <http://www.wired.com/threatlevel/2008/09/our-brand-is-cr/>
- [6] J. A. Vargas, *Major front in Virginia race is online*, Apr. 29, 2009. <http://www.washingtonpost.com/wp-dyn/content/article/2009/04/28/AR2009042803419.html>
- [7] Online Document, *FEC Internet Rulemaking - Background and FAQ*, Mar. 27, 2006. <http://www.fec.gov/members/weintraub/nprm/statement20060327.pdf>
- [8] P. Hamby, *McCain campaign pulls ads from some anti-Obama web sites*, Jul. 1, 2008. <http://www.cnn.com/2008/POLITICS/07/01/mccain.ads/>