2013 27th International Conference on Advanced Information Networking and Applications Workshops

A general purpose model for future prediction based on web search data: Predicting Greek and Spanish elections

Spyros E. Polykalas
TEI of the Ionian Islands
Dept. of Business Administration
Lixouri, Kefalonia, 28200 Greece
Tel:+30 26710 92861
s.polykalas@teiion.gr

George N. Prezerakos
TEI of Piraeus
Dept. of Electronic Computing Systems
P. Ralli & Thivon 250, Athens, 12244 Greece
Tel:+30 210 5381132
prezerak@teipir.gr

Agisilaos Konidaris
TEI of the Ionian Islands
Dept. of Business Administration
Lixouri, Kefalonia, 28200 Greece
Tel:+30 26710 92861
konidari @teiion.gr

Abstract— **Although real time data processing consumes a significant portion of computing resources worldwide, we are moving rapidly from the age of "real time" towards the era of "next time". The term "next time" characterizes the combination of real time data flows from collective sources with massive computing power with the aim of predicting the future. In other words, if you can compute "fast enough" using real time data sets then you can accurately predict what happens next. Central to this process is the Google Trends service that provides generalized statistics in regard to the popularity of web search terms submitted to Google. The paper combines the conclusions derived from other approaches to the prediction problem with Google Trends data in order to predict the outcome of six national elections races in two countries, Greece and Spain. The results of the proposed model reaffirm our hypothesis that web search terms popularity is directly related to voters decisions in both countries and thus can be used to predict the final outcome with great accuracy.**

Keywords-compenent:Google Insights, Google Trends, web search, prediction, elections.

I. INTRODUCTION

The abundance of processing power, realized on one hand by Moore's Law and on the other hand by "cloud" services supported by mega-corporations such as Google and Amazon, has led to the ability to analyze huge streams of data, the moment they occur. Such "real time" processing is nowadays crucial in several socioeconomic activities such as e.g. investment trading where a whole sector, known as super fast or high frequency trading, thrives on the ability to process stock exchange information almost instantly. However, recent articles in the press [10] indicate that "real-time" is not good enough and will be replaced sooner or later by "next-time". "Next-time" is a term that characterizes the ability to exploit huge amounts of computing power in conjunction with either historical or real time data in order to predict what will happen in the future.

Probably the best example of what 'next-time" really means is the Google Insights service by Google Inc[1]. This service allows any user in the Web to view and use statistics related to the popularity of any search term on the Web as long as it has been searched via the Google Search Engine.

The paper attempts to tackle the following problem: Given the right set of search terms, would it be possible to use such aggregated web statistics to predict election results? More specifically, the paper focuses on the prediction of the percentages of the two major parties in Greek and Spanish national elections by employing a simple but very effective model. The paper is structured as follows: Section 2 discusses the state of the art in predictions that exploit search and social networking data. Section 3 describes the model in question. Section 4 is concerned with the selection of search terms relevant to the Greek and Spanish national elections. Section 5 examines the four more recent elections in Greece and the two more recent in Spain and explains how the model can be used in order to predict the outcome. The paper concludes with Section 6.

II. RELEVANT WORK AND INNOVATION

As it has been mentioned earlier, web based prediction is a research area that engulfs several human activities. Google researchers have attempted predictions using search term popularity in a number of areas from home, automobile and retail sales to travel behavior [3]. Homes sales predictions have also been attempted by other researchers [11] while the work in [6] constitutes an interesting take on predicting flu epidemics before they actually emerge. The prediction of unemployment rates is another successful exercise that has been attempted both before the establishment of Google Insights for the U.S. job market [4] as well as afterwards in regards to Germany [1].

With respect to elections, an initial approach in [9] considers election predictions along with predictions in sports and economics. Google Insights was not around when that paper was written therefore the authors have performed normalizations directly on search results from Google. Consequently, in [8] Flickr has used as a source of data for the prediction of the winner in both the U.S. primaries and in the U.S. General Elections. In addition, Twitter has also been considered as an alternative data source for election results predictions [2], [12].

It should also be noted that, web based prediction, being a relatively recent research area, is subject to criticism especially from proponents of more traditional methods such as polls. In [13] is argued that there are strong limitations on the predictability power of a Google trend since it is difficult

[1] During the time when the paper was written, Insights was a stand-alone service. Currently Google has merged it with Trends at http://www.google.com/trends/

978-0-7695-4952-1/13 $26.00 © 2013 IEEE
DOI 10.1109/WAINA.2013.155

to determine the circumstances behind a user's search for the profile of a certain candidate and thus make a guess about that candidate's public image and why a user might be interested in him/her. In [5] it is stated that predictions of U.S. elections results based on Twitter perform only slightly better than pure chance; this conclusion does not include Flickr-based results in [8]. Also, research from the side of Yahoo Inc. [7] shows that in regard to retail sales, the predictive ability of web search based methods varies greatly with the product in question (music, films, video games).

Compared to existing research efforts, this paper is innovative in the following ways: (a) it proposes a model for election prediction based on web search data (b) it successfully applies the proposed model in regards to six election results prediction in two countries, Greece and Spain, (c) it accurately predicts not only the winner of each election but also the relative (normalized) percentages of the two major parties before election day, and (d) it does so by using a model which is computationally less intensive compared to the autoregressive equations used in most of the papers referenced already.

III. THE PREDICTION MODEL

The prediction model is mainly based on the assumption that, during the pre-elections periods, a high percentage of the actual voters of one party are searching for that party on Google, using words or and phrases related to the name of that party. We argue that this percentage is adequate to establish a relation between the search term popularity of a party, during the pre-election period, and the number of votes that this party will finally receive. Of course in reality the relation between the search term popularity and the final election results may differ between the various parties. For example, the profile of the potential voters of one party may be more internet friendly than the respective profile of the other party. In addition not every person that searches for a party, during the pre-election period, will vote for that party.

In order to reduce the noise in our predictions, generated by the conditions mentioned above, we calculate a factor that connects the web search interest for a party with its electoral percentage. In cases where the search behavior of the electorate of each party does not change drastically between consecutive elections, in particular compared with the search behavior of the electorate of the other parties, then one can calculate this factor for the pre-election period of a previous election race and use it to predict the results of a forthcoming election race. On the other hand in cases where major differentiations are observed in relation to the search behavior of the electorate of one party compared mainly with the relevant search behavior of the other parties, then the feedback from the previous election races is ignored and the predictions are solely based on the behavior of the electorate a few weeks before the date of the forthcoming elections.

The model operates on three different observation windows, which all of them have one month duration. The first window stops two weeks, the second one week and the third one day before the election date. For each such observation window, we query Google Insights for the web search interest values of the two major parties in each country.

More specifically, the initial dataset contains web interest values from Google Insights for the period from 2004 to 2012. The queries are initially carried out using the acronyms of the parties as search terms. This dataset is further divided into observation windows around the dates of elections. For each observation window, we process the respective web search interest values of the two major parties of the Spanish and Greek political scene. The web search interest values in our dataset go through the following steps:

Let $WI_{N,\ \text{party x, current elections}}$ be the web interest value for Party x during the current elections race, on the N^{th} day before the elections.

First we calculate the average web interest over a period of 30 days before the N^{th} day:

$$AWI_{\text{party x, current elections, period}} = \frac{1}{30}\sum_{N=1st\ day}^{N-30} WI_{N,\text{party x,current elections,period}}$$

We calculate the AWI for three different values of N: one day before the elections, seven days before the elections (one week) and fourteen days before the elections (two weeks).

For each AWI we normalize the average web interests of each party to a 100% to arrive at the Normalized Web Interest (NWI) for each party, by calculating first the Total Web Interest (TWI):

$$TWI_{\text{current elections, period}} = AWI_{\text{party y, current elections, period}} + AWI_{\text{party x, current elections, period}}$$

$$NWI_{\text{party x, current elections, period}} = \frac{AWI_{\text{party x, current elections, period}}}{TWI_{\text{current elections, period}}}$$

Finally we divide the NWI of each party with the respective percentage in the current elections, also normalized to 100%. The result, called Model Indicator (MI), expresses the relation between the web interest for one party with its actual elections results.

In order to predict the normalized percentage (NP) of each party we follow an algorithm based on the relation between the AWIs of the parties between the previous and the forthcoming elections. More specifically, we calculate the variances between the AWIs of the parties for the previous and the forthcoming elections. If the variances do not differ in absolute values more than 10%, the method A is used, otherwise method B is considered.

Method A

In order to predict the normalized percentage (NP) of each party in the next elections we are using the MI of the previous elections and the AWI of the next elections.

$$NP_{party\ x,\ current\ elections,\ period} = NWI_{party\ x,\ current\ elections,\ period} * MI_{party\ x,\ previous\ elections,\ period}$$

Method B

In method B we are not using historical data for our prediction, since the values of the AWI for the forthcoming elections indicate that the relation between the search behaviors of the parties' voters has dramatically changed

compared with the relevant relation for the previous elections. Therefore it is considered more appropriate for our predictions to use the current values of the WI, without feedback from the previous elections. This means that we use as the Normalized Percentage (NP) of each party in the next elections, the relevant NWI of the next elections.
In both methods, for each prediction of the normalized percentage, we compare the prediction of the model with the actual normalized percentage of the relevant election race. Since we are not in a position to know or to predict with accuracy, before the election, the actual sum of the percentages of the two major parties, we restrict our prediction to the normalized percentage of each party. Nevertheless, the prediction of the normalized percentage for each party represents among other things, the percentage difference between the actual percentages of the parties and therefore is a strong indication for the actual results of the election race.

IV. Selection of Search Items

The words that first come to mind as valid search terms are the political parties' acronyms. In Greece, the two major parties that consistently win the majority of Greek parliament during the elections are the Pan-Hellenic Socialist Movement (PASOK) and New Democracy (ND). Respectively in Spain the two major parties are the Spanish Socialist Workers' Party (SPOE) and People's Party (PP).
Afterwards, using Google Trends, we examine the evolution of web search interest for the search terms of each party for short periods before the election date. In Greece national elections took place on 7th March of 2004, 16th September 2007, 4th October of 2009, 6th May 2012 and 17th June 2012, while in Spain the dates of national elections were, 14th March of 2004, 9th March of 2008 and 20th November of 2011. We are focusing on the time period of a few weeks prior to election day, in order to examine whether the search term presents significant variation in this period. If yes then it is considered to be a valid search term for our model.

For the Greek election, due to the Greek alphabet, search interest for each term, has to be examined both in Latin and Greek letters, while in Spain the relevant terms for each party exist only in Latin characters.

Figure 1 displays the web search interest variation for the selected search terms (acronyms of the parties) related to the two major parties in Spain and Greece on a weekly basis as provided by Google Trend service. In the periods around the election races the web search interest for the two major parties, in each country, present peak values, while there are few other periods where some selected terms present also peak values. The latter peak values are related to internal procedures of these parties, such as the election of a new party leader.

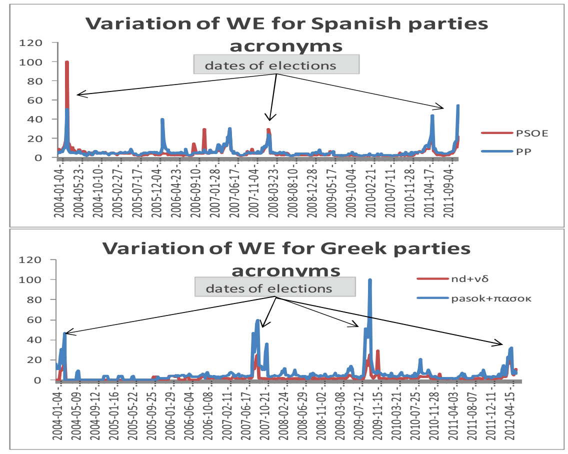

Figure 1. Web interest (WE) variation for parties's acronyms

In the next phase we examine more words and terms relevant to the names of each party, trying to include all words / terms which may be related to voters' willing to vote for a specific party in the forthcoming national elections. In order to examine whether a specific word or phrase should be included in the set of words / phrases for the relevant party, we apply the following rules: first we examine whether the variation of web interest presents peak values around the election days, similar to the web interest variation presented by the acronyms of the relevant parties and secondly we examine whether the values of web interest are at such a level that affects the relevant web interest of the party's acronyms. If both criteria are fulfilled then the word/phrase should be included in the set of word / phrase selected for a specific party. Applying the above rules in several runs we conclude that only for the Greek party of Nea Dimokratia a modification to the initial selected set of words is required. It should be also noted here that, in all cases, we restrict web search results at a national level. In our cases web searches have been restricted to Greece and Spain for Greek and Spanish elections respectively.

According to the rules stated previously, the final set of search terms in regard to the major parties of the Greek and Spanish political scenes is as follows: For Greece "Pasok, Πασοκ, νδ, nd, nea dimokratia, νεα δημοκρατια" , and for Spain "Psoe, Pp".

V. Election Study

A. Greek Elections

In the 2004 elections New Democracy (ND) was the winner. We execute the model using Google Insights data for periods around the date of the elections applying the mathematical types explained in Section 4. Table I shows the input parameters of the model.

TABLE I. GOOGLE INSIGHTS DATA PROCESSING FOR THE 2004 GREEK ELECTIONS

Greek Elections 2004					
		period	*PASOK*	*ND*	*Total*
Model Inputs	*WI*	two weeks	20.5	7.5	28.0
		one week	25.8	11.0	36.8
		one day	31.5	14.5	46.0
	Actual Election Results	actual	40.6%	45.4%	85.9%
		normalized	47.2%	52.8%	100.0%

In the 2007 elections New Democracy (ND) again was the winner. Consequently, we retrieve the Google Trends web interest values for periods around the elections date and we attempt to predict the outcome of the 2007 elections also taking into account the calculations that have been performed for 2004. The input parameters and the outcome of the model are presented in Table II.

TABLE II. MODEL INPUT, OUTPUT PARAMETERS FOR THE 2007 GREEK ELECTIONS

Greek Elections 2007					
		period	*PASOK*	*ND*	*Total*
Model Inputs	*WI*	two weeks	23.3	8.5	31.8
		one week	31.5	12.0	43.5
		one day	42.0	17.3	59.3
	Actual Election Results	actual	38.1%	41.8%	79.9%
		normalized	47.7%	52.3%	100.0%
Model Output (predictions)	*Normalized party's percentage*	two weeks	47.2%	52.8%	
		one week	50.1%	49.9%	
		one day	50.1%	49.9%	
	Model error	two weeks	-0.4%	0.4%	
		one week	2.4%	-2.4%	
		one day	2.4%	-2.4%	

For 2007, the model fails to predict the winner for periods of two and one week before the elections. However, one day before the elections, the model predicts the winner and the normalized percentages of the two major parties with accuracy more around 2,4%.

In the 2009 elections PASOK was the winner. We retrieve the Google Trends web interest values for a period around the election date and we attempt to predict the outcome of the elections also taking into account the relevant results from 2007. The model predicts the winner two as well as one week before the elections. In addition one day before the elections, the model predicts the winner and the normalized percentages of the two major parties with accuracy around 3,6%. The results of the execution cycle for the 2009 elections are summarized in Table III.

In the first elections of 2012 ND was the winner, while PASOK was the third party. As indicated in the Table IV the relation between the AWIs of PASOK and ND has changed. Indeed the AWIs of PASOK in the elections of 2009 were higher than the relevant AWIs of ND, while in elections of May 2012 the ND's AWIs were higher than AWIs of PASOK.

TABLE III. MODEL INPUT, OUTPUT PARAMETERS FOR THE 2009 GREEK ELECTIONS

Greek Elections 2009					
		period	*PASOK*	*ND*	*Total*
Model Inputs	*WI*	two weeks	29.3	8.8	38.0
		one week	38.8	12.0	50.8
		one day	48.5	16.0	64.5
	Actual Election Results	actual	43.9%	33.5%	77.4%
		normalized	56.7%	43.3%	100.0%
Model Output (predictions)	*Normalized party's percentage*	two weeks	52.7%	47.3%	
		one week	52.8%	47.2%	
		one day	53.1%	46.9%	
	Model error	two weeks	-4.1%	4.1%	
		one week	-3.9%	3.9%	
		one day	-3.6%	3.6%	

Therefore, for our predictions, method B was used which does not take into account historic data. The results of this execution cycle are summarized in Table IV. The relatively high percentage of model error even one day before the elections (7,3%) could be explained due to high uncertainty in the Greek political scene. For the first time in the last 40 years the first party was below 20% while for the first time since 1974, PASOK was below 14%.

TABLE IV. MODEL INPUT, OUTPUT PARAMETERS FOR THE MAY 2012 GREEK ELECTIONS

Greek Elections May 2012					
		period	*PASOK*	*ND*	*Total*
Model Inputs	*WI*	two weeks	9.3	5.5	14.8
		one week	11.3	12.0	23.3
		one day	15.3	16.3	31.5
	Actual Election Results	actual	13.2%	18.9%	32.0%
		normalized	41.1%	58.9%	100.0%
Model Output (predictions)	*Normalized party's percentage*	two weeks	62.7%	37.3%	
		one week	48.4%	51.6%	
		one day	48.4%	51.6%	
	Model error	two weeks	21.6%	- 21.6%	
		one week	7.2%	-7.2%	
		one day	7.3%	-7.3%	

As a viable government could not be established in Greece after the May 6 elections, the second elections of 2012 took place on the 17th of June. ND was the winner, while PASOK was again the third party. This time the model algorithm indicates the method A as the most appropriate. We retrieve the Google Insights web interest values for a period around the election date and we attempt to predict the outcome of the elections also taking into account the relevant results from the previous elections. The results of this running are summarized in the Table V.

TABLE V. MODEL INPUT, OUTPUT PARAMETERS FOR THE JUNE 2012 GREEK ELECTIONS

Greek Elections June 2012					
Model Inputs	***WI***	***period***	***PASOK***	***ND***	***Total***
		two weeks	13.3	12.8	26.0
		one week	6.8	10.5	17.3
		one day	7.3	11.8	19.0
	Actual Election Results	actual	12.3%	29.7%	41.9%
		normalized	29.3%	70.7%	100.0%
Model Output (predictions)	***Normalized party's percentage***	two weeks	30.2%	69.8%	
		one week	32.4%	67.6%	
		one day	31.5%	68.5%	
	Model error	two weeks	0.9%	-0.9%	
		one week	3.1%	-3.1%	
		one day	2.2%	-2.2%	

The model predicts the winner of the elections, while one day before the elections, the model predicts the normalized percentages of the two major parties with an accuracy around 2,2%.

Summarizing the results of the model for all Greek elections, it can be said that in all cases the model predicts the winner of the elections at least one day before the elections. Figure 3 displays predicted percentages one day before the elections alongside actual election results (both normalized to 100) for all Greek elections.

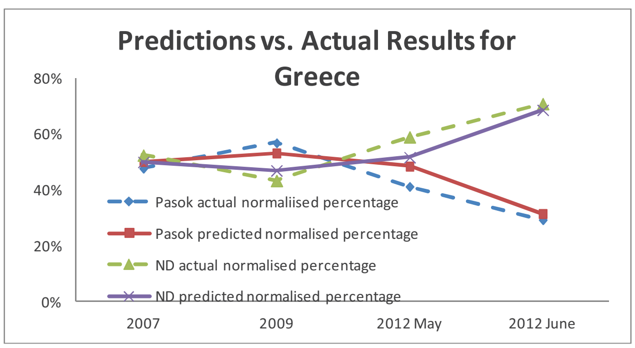

Figure 2. Greek Election results compared against predicted percentages

B. Spanish Elections

In this section we are applying the same methodology for Spanish national elections of 2008 and 2011. As input parameters, we have used the results of the previous elections of 2004 and 2008 respectively and the web search interest of the selected terms in a short period before the election date. Google Insights does not provide data earlier than 2004 so this year is once more our starting point.

In the 2004 elections PSOE was the winner. We execute the model using Google Insights data for a period around the date of the elections. The results are shown in the following table.

TABLE VI. GOOGLE INSIGHTS DATA PROCESSING FOR THE 2004 SPANISH ELECTIONS

Spanish Elections 2004					
Model Inputs		***period***	***PSOE***	***PP***	***Total***
	WI	***two weeks***	9.3	8.0	17.3
		one week	11.8	10.0	21.8
		one day	18.5	19.3	37.8
	Actual Election Results	actual	42.6%	37.7%	80.3%
		normalized	53.0%	47.0%	100.0%

In the 2008 elections PSOE was again the winner with a percentage 43,87% against PP with percentage 39.94%. Consequently, we retrieve the Google Trends web interest values for a period around the elections date and we attempt to predict the outcome of the 2008 elections taking into account these web search interest values as well as the calculations of the model for the previous elections race during 2004. The percentages of the two major parties are predicted one week before the elections dates with an accuracy better than 1.4%, while one day before the elections the error of our predictions is around 3,7% (Table VII).

TABLE VII. : MODEL INPUT, OUTPUT PARAMETERS FOR THE 2008 SPANISH ELECTIONS

Spanish Elections 2008					
Model Inputs		***period***	***PSOE***	***PP***	***Total***
	WI	two weeks	11.3	12.3	23.5
		one week	14.8	14.8	29.5
		one day	19.8	17.8	37.5
	Actual Election Results	actual	43.9%	39.9%	83.8%
		normalized	52.3%	47.7%	100.0%
Model Output (predictions)	***Normalized party's percentage***	two weeks	48.4%	51.5%	
		one week	50.9%	49.0%	
		one day	48.7%	51.4%	
	Model error	two weeks	-3.9%	3.9%	
		one week	-1.4%	1.4%	
		one day	-3.7%	3.7%	

In the 2011 elections, PP was the winner (Table VIII). Also for the 2011 Spanish elections, the model predicts the winner of the elections. The accuracy of the predictions is impressive, especially one day before the elections where the maximum error is less than 0,6%.

TABLE VIII. MODEL INPUT, OUTPUT PARAMETERS FOR THE 2011 SPANISH ELECTIONS

Spanish Elections 2011					
		period	*PSOE*	*PP*	*Total*
Model Inputs	*WI*	two weeks	5.5	9.0	14.5
		one week	7.3	11.5	18.8
		one day	9.0	14.5	23.5
	Actual Election Results	actual	28.7%	44.6%	83.8%
		normalized	39.2%	60.8%	100.0%
Model Output (predictions)	*Normalized party's percentage*	two weeks	34.7%	67.9%	
		one week	36.9%	64.4%	
		one day	38.5%	61.3%	
	Model error	two weeks	-5.4%	5.4%	
		one week	-2.7%	2.7%	
		one day	-0.6%	0.6%	

Figure 3 displays the predicted percentages one day before the elections alongside actual election results (both normalized to 100) for the two Spanish elections.

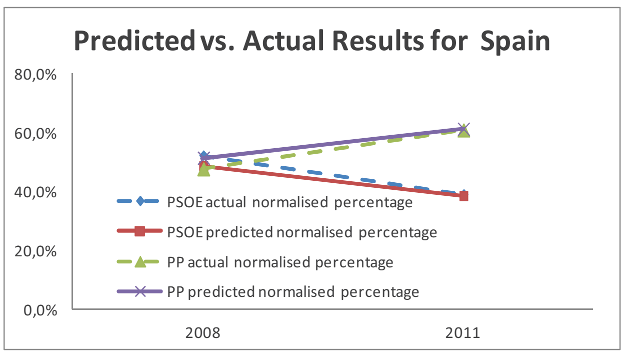

Figure 3. Spanish Election results for 2008 and 2011 compared against predicted percentages

VI. CONCLUSIONS

Recent research publications have shown that it is possible to predict future outcomes from currently evolving trends as long as these trends can be quantified using data from web search engines.

In the case of the Greek and Spanish elections, the paper shows that the normalized final result for the two major parties can be predicted with accuracy by combining data from Google Trends with a simple predictive model. A direct conclusion is that web search based predictions may soon become a viable alternative to traditional election polls when it comes to predicting the percentages of the participating parties. Of course, traditional polls still provide a richer result set since they can be used to approximate the socioeconomic profiles of the people who voted for each party. But at the end of the day, this is an issue mostly related to Google opening up their APIs to that kind of information as well. We suspect, although we cannot prove it, that someone with unrestricted access to Google's data set would come up with predictions that would be in par, if not superior, both in percentages and in demographic information to those provided by traditional polls.

Further research is required in order to eliminate the noise in our predictions generated in the cases where voters profiles (for the same party) become significantly different between consecutive election races, as well the noise generated by events related to political issues that occur in periods closed to election races and thus have major impact on the behavior of web search engine users. Nevertheless, we argue that there is always a strong correlation between the behaviors of web search engine users during the periods before election races and the actual election results.

REFERENCES

[1] Askitas N. and Zimmerman. K. F. 2009. Google Econometrics and Unemployment Forecasting. Applied Economics Quarterly (formerly: Konjunkturpolitik), Duncker & Humblot, Berlin, vol. 55(2), pages 107-120.

[2] Asur S. and Huberman B. A. 2010. Predicting the Future with Social Media. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01 (WI-IAT '10), Vol. 1. IEEE Computer Society. Washington, DC, USA, 492-499.

[3] Choi H. and Varian H. 2011. Predicting the present with google trends. Tech. report, Google, 2009. DOI=http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf., last accessed: June 24, 2011.

[4] Ettredge M., Gerdes J., and Karuga G. 2005.Using web-based search data to predict macroeconomic statistics. Commun. ACM 48, 11, pp. 87-92.

[5] Gayo-Avello D., Metaxas P. T. and Mustafaraj E. 2011. Limits of Electoral Predictions using Social Media Data. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, July 17-21, 2011.

[6] Ginsberg J., Mohebbi M., Patel R., Brammer L., Smolinski M., Brilliant L. 2009. Detecting influenza epidemics using search engine query data. Nature, vol. 457, pp. 1012-1014.

[7] Goel S,. Hofman J.M., Lahaie S., Pennock D. M., and Watts D. J. 2010. Predicting consumer behavior with Web search. In Proceedings of the National Academy of Sciences of the United States of America, Volume: 107, Issue: 41, Publisher: National Academy of Sciences, pp: 17486-17490.

[8] Jin X., Gallagher A., Cao L., Luo J. and Han J. 2010. The Wisdom of Social Multimedia: Using Flickr for Prediction and Forecast. In Proceedings of 2010 ACM Multimedia Int. Conference.

[9] Pion S. and Hamel L. 2007. The Internet Democracy: A Predictive Model Based on Web Text Mining. In Proceedings of DMIN 2007, pp. 292—300.

[10] Wired.co.uk staff .2012. Forget real time: 'next time' is far more disruptive. Wired.co.uk, 2012, DOI=http://www.wired.co.uk/news/archive/2012-02/20/forget-real-time, last accessed: March 2nd, 2012.

[11] Wu L. and Brynjolfsson E. 2009. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. NBER Conference Technological Progress & Productivity Measurement.

[12] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, Isabell M. Welpe: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM 2010.

[13] On the predictability of the U.S. elections through search volume activity, Catherine Lui, Panagiotis T. Metaxas, Eni Mustafaraj, IADIS, 2011