ASSESSING PUBLIC OPINION ON SOLAR ENERGY TOPICS USING GOOGLE TRENDS

Alexandra PERJU-MITRAN [1]
Raul-Augustin MITRAN [2*]

ABSTRACT

Public interest towards environmental and green energy topics is a vital part shaping both political action and consumer behavior towards these topics. The interest shown by the public towards various topics can be gauged by the frequency of online search queries. This "big data" approach can be used in topics ranging from epidemiologic to environmental or financial topics. Google provides access to all search queries performed stating from 2004 using its online search engine through the Google Trends service. We present a study of multiple search keywords associated with solar energy in the 2004-2017 time span. Our results show a marked difference between keywords denoting large scale or small scale solar installations, as well as a shifting of public interest from the former to the latter, which can be associated with decreasing prices for photovoltaic solar panels. The total search frequency related to solar energy topics shows a maximum in spring 2008, followed by a decline up to 2014 and an increase in the last years.

KEYWORDS: *Google Trends; solar energy; big data; public interest; consumer behavior*

1. INTRODUCTION

Global warming is arguably the most impressive challenge facing humanity in the 21st century. An important step in addressing the anthropologic climate chance is transitioning from fossil fuel to renewable energy, such as solar and wind. Public opinion is an important parameter factoring in current and future policies, [1] and it is expected to influence renewable energy topics. The public interest is influenced by both online and offline media [2] and it can be gauged by the relative frequency of specific keywords used in Internet search engines. A large number of studies so far have used publicly available Google search data as a measure of public interest towards various topics, such as flu outbreaks, [3] environmental concern,[4, 5] traffic and housing conditions, [6] finance [7] etc. Moreover, Google search data was shown to have the same accuracy as surveys in assessing the public interest towards a specific topic. [6]

[1] Assistant Professor PhD, Romanian-American University, Faculty of European Economic Studies, Bucharest, alexandraperju@gmail.com
[2*] corresponding author. Research scientist, 3rd degree, "Ilie Murgulescu" Institute of Physical Chemistry, Romanian Academy, Bucharest, raul.mitran@gmail.com

One of the most important sources of renewable energy sources is represented by solar power. So far, the interest towards solar power has not been assessed using Google search engine data. It is important to note that public interest towards most environmental concerns, with the exception of climate change, is declining. [4] Herein the public interest towards both general and specific solar energy as well as towards large scale or small scale solar installation is evaluated using a series of eleven search engine keywords, in the 2004-2017 time range. This study deals only with Google searches performed in English, throughout the world in the investigated time span. The results are discussed and they show a shift from passive interest in general and large scale solar projects towards active interest in specific and small scale installations.

Understanding the functionalities of potential consumers' online search process allows identification of new models, interaction between buyers and sellers, and new functionalities for organizations and companies. Studies on search data and online consumer behavior can be found in a variety of publications, covering areas such as information technology, information systems, healthcare, psychology and marketing.

2. METHODS

The query search data is provided by Google as a time series for each keyword, starting from 2004. The Google data can be freely downloaded from the Google Trends website (trends.google.com). Each data point is scaled with respect to total searches performed for a time range and location and normalized to the highest data point in a time series. [4] the data is typically provided to cover 1 month time range and expressed between 0 and 100, with 0 representing less searches than a given threshold value.

In order to account for different use of synonymous terms, it is essential to compare an array of related terms. [5] A number of 11 search keywords related to solar energy were selected ("Solar power", "Solar energy", "Solar panel", "Solar cell", "Concentrated solar", "Solar thermal", "Photovoltaic", "Solar panels price", "Solar cost", "Solar roof", "Rooftop solar"). In addition, a baseline query was performed by using the term "the", the most commonly used word in English. The baseline can be used to control for artificial trends arising from changes in search user number and language. [6] The individual search keywords were selected from the topic of solar energy, using the "Related queries" feature of Google trends. Keywords with insufficient search volume data were discarded.

3. RESULTS AND DISCUSSION

The Google Trends data is available as a time series for each individual keyword (Fig. 1). The values are proportional to the frequency of each keyword in the total searches performed in the specified time range. Google normalizes each time series in the 0 -100 range, so that 100 corresponds to the highest value of the time series. This methodology makes the data values impervious to changes in total search volume or number of search users. However, the time series values can be artificially affected by changes in user search behavior or interruptive events. [6] In order to investigate such changes, the keyword "the", being the most commonly used word in English was used as a baseline (Fig. 1 A). The baseline shows a slight decrease from 2004 (92.0 ± 6.2 %) to 2017 (86.3 ± 3.1 %), with a minimum in 2007 (78.3 ± 2.8 %). This behavior characterizes changes in

search engine user behavior, such as use of English language. Two trends can be noticed with respect to individual keyword time series in Fig. 1. Six out of 11 keywords ("Solar power", "Solar energy", "Solar cell", "Concentrated solar", "Solar thermal", "Photovoltaic") show a maximum in 2008 followed by a decreasing values up to 2018 (Fig. 1 B, C, E, F, G, H). In contrast, the rest of the five keywords ("Solar panel", "Solar panels price", "Solar cost", "Solar roof", "Rooftop solar") show a continuously increase from 2004 to 2018 (Fig. 1 D, I, J, K, L). The 2008 maximum might be related to the 2008 financial crisis, as it precedes the bankruptcy of Lehman Brothers collapse in September 2008. The "Solar roof" keyword is a particular case, showing two large spikes in search data in May 2017 and November 2016. This can be easily explained by the introduction of the Tesla "Solar roof" product in 28 October 2016. [8] Another particular case is the "Concentrated solar" keyword. Even though it shows positive gains between 2004 and 2017, the frequency of this keyword has been steadily decreasing since 2008, the positive change being associated in this case with a very low search data in 2004.

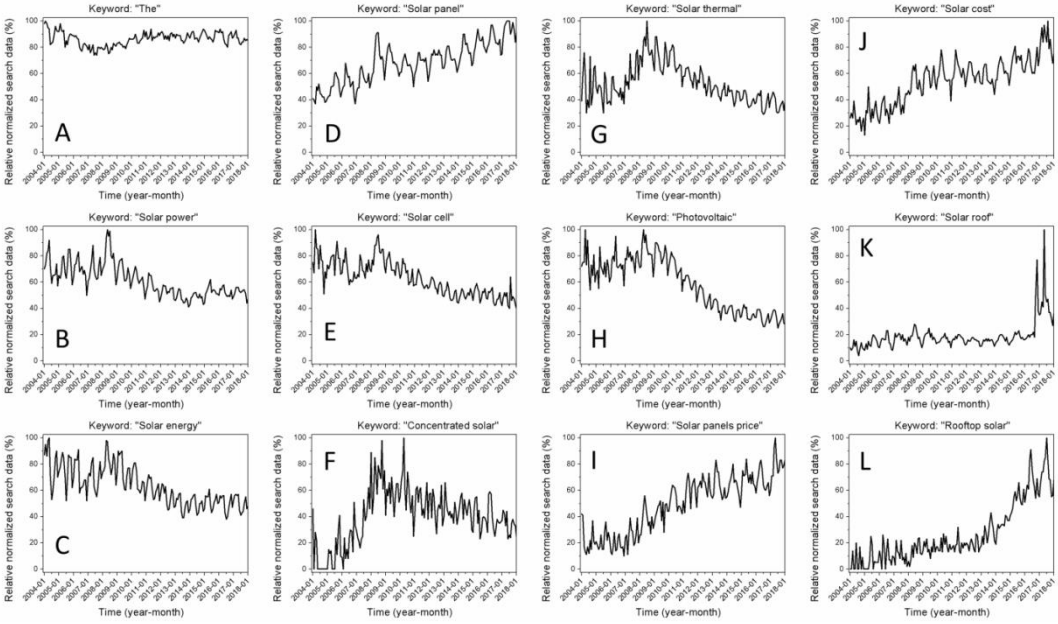

Figure 1. Normalized Google search data

It can be noticed that the magnitude of the change for the keywords with decreasing frequency is smaller than for the keywords with increasing frequency between 2004 and 2017 (Table 1). Indeed, the sum of the normalized search data change between 2017 and 2004 for all 11 keywords is positive and equal to 161 ± 46. It is worth noting that Google Trends search values are not absolute values and therefore the sum of the change rate for all keywords cannot be taken as a measure of overall public interest towards the investigated issue.

Table 1. Average normalized Google search data for each keyword in 2004, 2008, 2012, 2017 and search data average difference between 2017 and 2004. Values presented as average ± standard deviation for each year.

Keyword	*2004*	*2008*	*2012*	*2017*	*2004-2017 Difference**
The	92.0 ± 6.2	79.5 ± 2.5	87.4 ± 2.5	86.3 ± 3.1	-5.7 ± 3.8
Solar power	71.6 ± 10.4	81.3 ± 13.2	47.7 ± 4.1	52.4 ± 3.4	-19.2 ± 8.5
Solar energy	80.3 ± 15.5	80.9 ± 10.0	48.8 ± 6.5	50.2 ± 6.1	-30.1 ± 9.8
Solar panel	43.1 ± 4.7	72.1 ± 13.3	71.4 ± 6.2	92.9 ± 7.8	49.8 ± 6.8
Solar cell	75.5 ± 13.1	82.6 ± 8.7	50.9 ± 4.6	47.8 ± 6.4	-27.7 ± 13.6
Concentrated solar	8.0 ± 15.4	71.1 ± 16.0	44.4 ± 8.7	34.5 ± 7.4	26.5 ± 15.5
Solar thermal	51.6 ± 16.2	78.3 ± 9.4	46.8 ± 4.2	37.3 ± 5.5	-14.3 ± 15.8
Photovoltaic	73.8 ± 12.8	82.8 ± 10.5	41.0 ± 4.9	32.8 ± 4.4	-41.1 ± 10.4
Solar panels price	22.3 ± 11.3	37.3 ± 9.8	63.6 ± 10.2	79.1 ± 10.6	56.8 ± 19.5
Solar cost	26.0 ± 5.6	53.2 ± 8.7	55.6 ± 7.0	84.5 ± 9.9	58.5 ± 11.6
Solar roof	10.3±3.4	19.1±5.2	14.7±2.1	44.3 ± 18.7	34.1 ± 18.8
Rooftop solar	5.0±7.8	13.5±7.0	27.9±7.8	72.2±13.8	67.2 ± 15.4

* Computed as $\frac{\sum_{m=1}^{12} f(2017,m)-f(2004,m)}{12}$, where *m* represents the month (1-12) and *f*(*y*, *m*) represents the normalized Google search data for year *y* and month *m*.

In order to gleam the relationships between the keywords, a correlation study was performed, by computing correlation coefficients for each keyword pair (Table 2). The coefficients take values between -1 and 1, with values close to ±1 indicating either positive or negative correlation and values close to 0 indicating no correlation. Two sets of correlated keywords, with coefficients above 0.80 can be noticed: the decreasing frequency set ("Solar power", "Solar energy", "Solar cell", "Photovoltaic") and the increasing frequency keywords ("Solar panel", "Solar panels price", "Solar cost", "Rooftop solar"). "Concentrated solar" and "Solar thermal" are correlated with the decreasing frequency while "Solar roof" is correlated with the increasing frequency set, with lower coefficients, in the 0.50-0.70 range. The correlation study confirms the existence of the two sets of keywords, denoted herein decreasing and increasing, based on their time series behavior.

Table 2. Correlation coefficients of the investigated keywords. Values above 0.8 are bolded.

Keyword	*Solar power*	*Solar energy*	*Solar panel*	*Solar cell*	*Concentrated solar*	*Solar thermal*	*Photo-voltaic*	*Solar panels price*	*Solar cost*	*Solar roof*	*Rooftop solar*
Solar power	1.00										
Solar energy	**0.88**	1.00									
Solar panel	-0.30	-0.46	1.00								
Solar cell	**0.88**	**0.84**	-0.47	1.00							
Concentrated solar	0.01	0.02	0.38	0.05	1.00						
Solar thermal	0.58	0.62	-0.15	0.65	0.54	1.00					
Photovoltaic	**0.84**	**0.86**	-0.55	**0.86**	0.09	0.71	1.00				
Solar panels price	-0.56	-0.57	***0.85***	-0.64	0.37	-0.22	-0.70	1.00			
Solar cost	-0.37	-0.41	***0.91***	-0.47	0.48	-0.07	-0.52	***0.88***	1.00		
Solar roof	-0.12	-0.22	0.59	-0.29	0.06	-0.19	-0.31	0.47	0.60	1.00	
Rooftop solar	-0.50	-0.60	***0.80***	-0.65	0.05	-0.50	-0.74	0.73	0.74	*0.63*	1.00

The existence of the two sets of correlated keywords related to solar energy can be explained by a combination of two mechanisms. In some areas, such as computers, the search results data follows the market share of specific products [9, 10] and shifts in vocabulary from general to specific terms. [4, 11] The keywords in the decreasing frequency set can be viewed as general terms (*i.e.* "solar energy", "solar power", "solar cells"), while the increasing frequency set contains more specific or refined terms (*i.e.* "solar panels", "Rooftop solar"). The trend of replacing general terms with specific ones may indicate an increase in public education towards solar energy issues.

A better delimitation of the two sets is scale. Thus, the keywords in the increasing set are associated with small-scale solar installation, such as rooftop solar, pertaining to individuals or small-medium enterprises. The decreasing set in contrast describes large scale solar installations, the purview of governments or large corporation. The best argument for a delimitation between small and large scale of the two sets is the "Concentrated solar" keyword, which denotes a specific solar technology which is currently available only at large scale. No concentrated solar power plants were constructed between 1991-2005, while the total installed concentrated solar power (CSP) increased from 355 MW in 2005 to 484 MW in 2008, 2553 MW in 2012 and 4815 MW in 2016. [12, 13] The remarkable increase in CSP is not reflected in the Google search data for this keyword, supporting the hypothesis that the two sets of decreasing and increasing frequency sets are related to a shift in public interest from large scale solar to small-scale installations.

Even though the search data is both relative and normalized, Google provides an option to compare up to five keywords using the same scale. In this scale, the 100 value is represented by the maximum data point present in the five keywords. It is therefore possible to estimate the general search data volume for all the investigated keywords, by always keeping a given time series as a reference and comparing it to four other keywords. Three time series were constructed by keeping the "solar energy" keyword as a reference. Three aggregate time series were obtained, which represent the six decreasing keywords, the five increasing terms and their overall sum. These are denoted "Increasing", "Decreasing" and "Overall", respectively (Fig. 2). The decreasing aggregate shows the same behavior as the individual time series, having a maximum in spring 2008, followed by a sharp decline until 2014 and relatively steady levels since then. The increasing aggregate series presents a linear increase from 2004 to 2017. However, the relative magnitude of the increasing aggregate is lower than that of the decreasing series. The overall data for the 11 investigated keywords is similar to that of the decreasing series, since the decreasing terms have higher search volume than the increasing keywords. Nonetheless, the influence of the increasing terms can be seen as an upward slope from 2014 to 2017. The overall search volume related to the topic of solar energy topic is still below the maximum 2008 levels.

The overall interest gauged by Google search data does not follow the total worldwide installed solar capacity nor the rate of new yearly installations (Fig. 2 B). [14] Both total installed and new yearly installed solar energy have been increasing at an accelerated rate since 2004, even during the 2008 financial crisis. The linear increase for the keywords related to small-scale solar installations seems to follow the decrease in PV price. [15] For

example, PV prices in Europe have decreased linearly from 1999 to 2012, with the latter being ~50% of the former. [15]

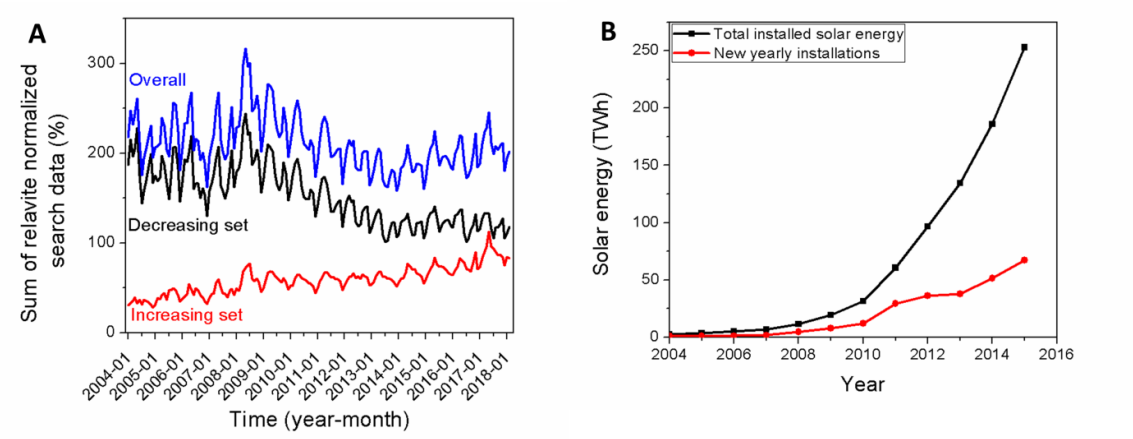

Figure 2. A) Aggregate search data for the six decreasing and five increasing keywords, as well as their overall sum. Data is normalized to the highest value of the "solar energy" time series, taken as 100. B) World installed solar energy and new yearly installations. Data from [14]

The Google search volume for individual keywords also contains localization data. There is a great deal of variance between the investigated keywords in terms of interest by region as well as number of regions with more than 1% searches with respect to the top country. Some keywords elicit interest in a large number of countries (i.e. "Solar energy" with 64 out of 251 geographic regions having more than 1% of the maximum), while others have Google searches coming from a small number of regions (i.e. "Concentrated solar" with only 1 region). The search distribution versus geographical region is presented in Table 3 and Table 4 for the keywords belonging to the decreasing and increasing sets, respectively.

The data is biased towards regions and countries which use Google search engine with English words, as well as countries in which English is a major language. [16] The data is expected to also present a marked socio-economic bias, as countries with poor internet penetration or low English literacy are underrepresented. These biases might explain the leading position of Australia in overall searches towards solar energy topics. Australia occupies a position in top three regions for three out of six decreasing keywords and all five increasing keywords. Nonetheless, the geographic distribution of solar energy searches shows a global interest towards this topic. It is interesting that many developing countries have top positions for individual keywords: Jamaica, India, Mauritius, Ethiopia, Sri Lanka, Cyprus, Nigeria, Pakistan, Thailand etc. This fact can be explained by increased interest in developing countries for solar energy as a part of their energy strategy.

Table 3. Geographical distribution for the search keywords from the "Decreasing" set. Only the top 25 regions are presented, with the numbers in the table header representing the number of regions with at least 1% searches with respect to maximum

Solar energy (64)		**Solar power (58)**		**Solar cell (57)**		**Concentrated solar (1)**		**Solar thermal (25)**		**Photovoltaic (66)**	
Jamaica	100	Australia	100	South Korea	100	United States	100	India	100	Mauritius	100
Australia	73	South Africa	69	Ethiopia	64			Australia	99	Malta	83
Ethiopia	73	Sri Lanka	65	Thailand	53			United Kingdom	86	Cyprus	71
Trinidad & Tobago	72	India	63	Indonesia	51			Ireland	74	Greece	32
India	69	New Zealand	59	Singapore	51			United Arab Emirates	61	Lebanon	31
South Africa	66	Nigeria	55	India	47			South Africa	59	South Korea	28
Ghana	64	Philippines	47	Taiwan	47			Pakistan	58	Singapore	26
Nigeria	64	Ghana	42	Pakistan	44			Singapore	55	India	22
Nepal	61	Kenya	42	Bangladesh	39			Canada	52	South Africa	22
Pakistan	51	United Arab Emirates	41	Hong Kong	32			Philippines	50	Puerto Rico	21
United Arab Emirates	51	United States	40	Sri Lanka	32			United States	50	Australia	20
Oman	50	Canada	36	Nigeria	30			Malaysia	39	Hong Kong	20
Kenya	45	Qatar	34	Malaysia	25			South Korea	37	Ireland	19
Philippines	44	Pakistan	33	Australia	22			Iran	28	Jordan	19
Singapore	40	Singapore	33	Iran	21			Israel	24	Malaysia	19
United States	38	United Kingdom	31	Egypt	20			Thailand	18	Nigeria	19
Lebanon	36	Puerto Rico	29	Philippines	20			Spain	17	Kenya	17
Canada	34	Bangladesh	27	China	19			Netherlands	14	Tunisia	17
Qatar	34	Hong Kong	21	United Arab Emirates	18			Germany	13	United Kingdom	17
New Zealand	31	Malaysia	21	United States	18			Romania	12	Taiwan	16
Puerto Rico	31	Ireland	19	Kenya	17			Italy	11	United Arab Emirates	16
Bangladesh	27	Saudi Arabia	11	Iraq	15			France	7	United States	16
Hong Kong	27	Norway	10	South Africa	15			Indonesia	7	Iran	15
United Kingdom	25	South Korea	10	Canada	14			Mexico	7	New Zealand	15
Jordan	24	China	8	Jordan	14			Brazil	3	Algeria	14

Table 4. Geographical distribution for the search keywords from the "Increasing" set. Only the top 25 regions are presented, with the numbers in the table header representing the number of regions with at least 1% searches with respect to maximum

Solar panel (66)		**Solar panels price (5)**		**Solar cost (8)**		**Rooftop solar (7)**		**Solar roof (42)**	
Australia	100	Pakistan	100	Australia	100	India	100	Australia	100
Nigeria	95	Australia	66	India	71	Thailand	53	United States	60
Pakistan	91	India	37	United States	66	Australia	22	Canada	52
Philippines	85	United Kingdom	17	South Africa	51	Canada	17	United Kingdom	52
South Africa	73	United States	13	Canada	43	United States	16	New Zealand	50
Nepal	71			United Kingdom	41	United Kingdom	5	South Africa	46
India	66			Pakistan	32	Germany	2	Sri Lanka	30
Ghana	61			Philippines	20			Ireland	29
Chile	59							India	25
Kenya	59							Singapore	23
Puerto Rico	58							Thailand	21
Sri Lanka	56							United Arab Emirates	21
Panama	55							Nigeria	19
Singapore	55							Philippines	19
New Zealand	50							Malaysia	16
United States	48							Norway	15
Canada	47							Denmark	13
United Arab Emirates	47							Hong Kong	13
United Kingdom	47							Netherlands	12
Bangladesh	44							Pakistan	12
Colombia	41							Sweden	12
Guatemala	41							South Korea	8
Malaysia	41							Switzerland	8
Peru	40							Belgium	7
Costa Rica	39							Czechia	7

4. CONCLUSIONS

The public opinion towards solar energy topics was investigated using publicly available Google search data between 2004 and 2017. Eleven English search terms were selected and compared with a baseline word. All keywords show one of two separate behaviors. On one hand, six keywords show a declining use with time, presenting a maximum in spring 2008. The other five show a linear increase with time. The search terms with increasing use have lower search volume than the six keywords with a decreasing behavior. The overall search volume towards solar energy topics presents a maximum in 2008, followed by a decline up to 2014 and increase from 2014 to 2017.

The decreasing search keywords are either general or they describe large scale solar installations ("Solar power", "Solar energy", "Solar cell", "Concentrated solar", "Solar

thermal", "Photovoltaic", while the increasing terms are mainly associated with specific solar issues, such as cost and technologies, with the exception of "solar panel" ("Solar panel", "Solar panels price", "Solar cost", "Solar roof", "Rooftop solar"). The Google search behavior can be explained through two main factors: an increase in public education towards solar energy topic, leading to a change from general to specific terms and a shift in public perception from large scale solar installation towards small scale or individual solar projects. The results show that a shift in public opinion from passive interest towards large scale solar projects to an active role in small scale installations such as rooftop solar panels is occurring. This shift is not correlated with the worldwide installed solar capacity or yearly changes in such, but rather with the decline of photovoltaic solar power price, which has followed a similar linear trend with time. It can be therefor noticed that the individual solar power consumer appears to be highly price sensitive.

Even though the search data is biased towards English speaking regions with good internet penetration and literacy, the geographic localization shows high interest towards solar energy from both developed and developing countries.

Under the dual, theoretical and practical incidence of the present study, and under the recent search engine marketing developments, the results enrich the scientific field, offering the possibility to continue research or integrate findings into company decision-making systems for marketing communication. Given that in the online environment consumers have numerous alternatives and sources of information, emphasis should be placed on a more thorough knowledge of perceptions, then on the choice of the communication context and the content of the communication.

REFERENCES

[1] B.I. Page, R.Y. Shapiro, Effects of public opinion on policy, American political science review, 77 (1983) 175-190.

[2] C. Warren, Risk and the sacred: environment, media and public opinion in Bali, Oceania, 82 (2012) 294-307.

[3] H.A. Carneiro, E. Mylonakis, Google trends: a web-based tool for real-time surveillance of disease outbreaks, Clinical infectious diseases, 49 (2009) 1557-1564.

[4] M.L. McCallum, G.W. Bury, Public interest in the environment is falling: a response to Ficetola (2013), Biodiversity and Conservation, 23 (2014) 1057-1062.

[5] M.L. Mccallum, G.W. Bury, Google search patterns suggest declining interest in the environment, Biodiversity and Conservation, 22 (2013) 1355-1367.

[6] J. Zhu, X. Wang, J. Qin, L. Wu, Assessing public opinion trends based on user search queries: validity, reliability, and practicality, in: The Annual Conf. of the World Association for Public Opinion Research, 2012, pp. 1-7.

[7] L. Kristoufek, Power-law correlations in finance-related Google searches, and their cross-correlations with volatility and traded volume: Evidence from the Dow Jones Industrial components, Physica A: Statistical Mechanics and its Applications, 428 (2015) 194-205.

[8] L.A. Shuster, Tesla's Solar Roof Tiles Go on Sale, Civil Engineering—ASCE, 87 (2017) 26-26.

[9] M.L. McCallum, G.W. Bury, Google search patterns suggest declining interest in the environment, Biodiversity and Conservation, 22 (2013) 1355-1367.

[10] A. Heath, Mac market share continues to rise while PC shipments decline. Cult of Mac. January 11, in, 2013.

[11] T. Levitzky-Aviad, B. Laufer, Lexical properties in the writing of foreign language learners over eight years of study: Single words and collocations, C. Bardel, C. Lindqvist, & B. Laufer (Eds.) L, 2 (2013) 127-148.

[12] C. Lins, E. Musolino, K. Petrichenko, W. Rickerson, J. Sawin, K. Seyboth, J. Skeen, B. Sovacool, F. Sverrisson, L. Williamson, Renewables 2014 Global Status Report, (2014).

[13] J.L. Sawin, F. Sverrisson, K. Seyboth, R. Adib, H.E. Murdock, C. Lins, I. Edwards, M. Hullin, L.H. Nguyen, S.S. Prillianto, Renewables 2017 Global Status Report, (2017).

[14] B. Dudley, BP statistical review of world energy 2015, in, London, UK, 2015.

[15] C. Candelise, M. Winskel, R.J.K. Gross, The dynamics of solar PV costs and prices as a challenge for technology forecasting, Renewable and Sustainable Energy Reviews, 26 (2013) 96-107.

[16] A.F. Al-Eroud, M.A. Al-Ramahi, M.N. Al-Kabi, I.M. Alsmadi, E.M. Al-Shawakfa, Evaluating Google queries based on language preferences, Journal of Information Science, 37 (2011) 282-292.