Submitted: 15 February, 2018
Revised: 29 May, 2018
Accepted: 5 June, 2018

ANALYSIS OF GOVERNMENT AIDED UNIVERSITY IN KOLKATA: A WEBOMETRIC STUDY THROUGH GOOGLE SEARCH ENGINE

Bidhan Dolai*

Abstract *This article tries to calculate the web impact factor, based on Google search engine. Thirteen Government aided universities are introduced in this study. The Inlink web impact factor analysis, WISER value calculation, making the rank of these universities and calculation of correlation among the inlink web impact factor ranks with WISER rank are also introduced here. The area of this study is focused on Kolkata region only.*

Keywords: *Webometric, Kolkata, Government Aided University, Google Search Engine*

INTRODUCTION

The Internet and web technology has created a new and unprecedented environment to the government, business, educational institutions and individuals enabling them to webcast any information using multimedia tools. The World Wide Web (WWW) can provide information about anything, anyone, and anywhere.

The journey was begun in the year of 1955 at Stanford University by two persons, Larry Page (Lawrence Edward Page) and Sergey Brin (Sergey Mikhaylovich Brin). In 1995 Larry Page was considering a transfer to Stanford University's graduate program in the computer science. Lawrence Edward Page was an American computer scientist. He was a Chief executive officer (CEO) of Google parent company Alphabet INC. In the year of 1973, 23 march Lawrence Edward Page born at Palo Alto, California in the country of United State of America. Sergey Mikhaylovich Brain (Sergy Brain), A Russian origin scientist brought up in the United State of America, was a resident of Los Altos, California, United State of America. Larry Page and Sergey Brin together launched the Google Company and Google search engine, which had become most wildly, used a web-based search engine in the year of 1996.

Google developed a web-based search engine and it is the most used search engine in the world. Google is now available in 123 languages. Though in the year of 1995 it was under the project of Stanford University but was making the development of this search engine, it took a part in the world number one search engine which is most widely used.

For an academic institution, a librarian must look forward to building a library website which has targeted audiences. Most of the cases for an academic library as well as university library targeted audiences are researchers, faculties, students, and staff. In today's world, if we try to find the peoples who are not aware of internet and WWW (World Wild Web), it will be really rarest fact. The invention of the computer in the year of 1822 by Charles Babbage and rapid changes of the internet from the year of 1983 by Tim Berners-Lee, access to information is easy now a day. It is possible only through WWW era. The growth of website now a day is increased rapidly. The effect of globalization website is taking part in every matter. The remarkable increase in web service, as well as Information and Communication technologies, have provided opportunities to communicate with the targeted user to get information in better and easy way. Although Service through the website have some complication though it plays a key impact on the way people live, work and play world wild. We know all that what is a website actually. It is simple to define as, a set of related WebPages containing in a single domain. We can see the different website in the web environment; generally, these are a personal website, commercial website, a government website and nongovernmental website.

Let, see about search engine, Google is a search engine owned by Google, Inc. whose mission statement is to "organize the world's information and make it universally accessible and useful". The largest search engine on the web, Google receives over 200 million queries each day through its various services.

* University of Calcutta, Kolkata, West Bengal, India. Email: bidhandolai93@gmail.com

Another Here Inlink means, To Web page W, an Inlink is a URL of another Web page which contains a link pointing to W. Web Impact factor, it is a measure of the influence of a site across the entire Web calculated according to the number of links from other sites.

LITERATURE REVIEW

ST. Dina, Agusten and SKom, Winarti (2013) describe in their article 'Analysis on Backlinks and Page Rank of Automotive Company Websites Using Crawler Tool' so that company can easily understand their website in different search engine environment. This analytical article helps the companies for better marketing through their website. In this article, the author uses a tool named as SEO SpyGlass. By using this tool, the company can improve its marketing strategy worldwide. Another side 'Calculating web impact factor for University website of Pakistan' (2015), this article extracts from emerald insight. Arif Khan and Haroon Idres, the author of this manuscript analysis of link and calculate impact factor. Besides, this study elucidates to calculate the revised web impact factor of top ten universities website in Pakistan. Rank calculation of this university is also included. It is more important to compare web impact with other Pakistani University other than this studied university which is included in this article. Google and Bing are used to collect data. The reason for using two search engines for cross verifies the result is indexing techniques applied by the search engine are different in terms of technology and approach (Lee, 2008; Murphy, 2008). Computation of a maximum number of in links (external backlink) that refers to this study. The WIF usually distinct as the percentage connecting a total number of linked pages (self-link and inlink) and number of WebPages published on the website which are indexed by a search engine (not all the pages of the website). The author Arif Khan and Haroon Idrees talk about different search strategy for data collection to calculate R-WIF.

Fereshteh Didegah and Mohammad Amin Erfanmanesh (2010), introduce a study aimed at Malaysian Public University' websites using correspondence analysis method. The author of this study discusses two research questions first one, which Universities' website fare better than others based on evaluation criteria. Another, using correspondence analysis, determine groups in which Malaysian universities can categorize into and what are the attributes of each group? Traffic rank, average number of pages viewed, time spent on the site per user, number of links received from other websites and percentages of Malaysian and foreign visitors are examined in this study. An investigation is made through Alexa Internet data, used to collect data. Nineteen Malaysian public universities are taken in this study. The present research findings provide an evolution of Malaysian Public universities' websites status in term of their functions and performance on the web. Although, all universities have not performed well in all criteria, each university site has benefits and advantages. Website administrator should pay more attention to the service provided in order to increase user satisfaction. One way to do this is to introduce their websites to other sites on the web in order to gain more links. Connectivity plays an important role in recognizing websites in extended virtual environments. Another way to enhance website ranking is to provide services that attract the international visitor.

Universal Journal of education research published an article (2016), 'webometrics ranking in the context of accessibility of higher education' written by Margarita, Bershadskaya, Yalia voznesenskaya and Olga Karpenko. This study is conducted for a comparative assessment of the development of mass higher education in the regions and the countries on the basis of the result of webometrics. International Journal on recent and innovation trends in computing and communication published (2014) an article named as 'Analysis of web Crawling Algorithms' by Rashmi Janbandhu, Prashant, dahiwale, M.M. Raghuwanshi. This discusses web crawling algorithms. They contribute their writing to say that the web is a huge and titanic collection of data and popularity of web is increasing now a day. The authors discuss their relevant pages for search topic; the results are still huge to be explored. The performance of a crawler depends mostly on the opulence of link and the specific topic being searched. This paper reviews the researcher on the web crawling algorithms used for searching.

OBJECTIVES

The objectives of this research are to identify:

- Inlink web impact factor analysis of Government aided University website in Kolkata.
- Ranking of Government aided University website in Kolkata according to Inlink web impact factors in Google search engine.
- Calculation of WISER value of this University and making a rank depending on WISER value.
- Calculate rank correlation among Inlink web impact factors and WISER ranking.

METHODOLOGY

The validity of any research depends on the systematic method of the data collection and analyzing the data in a logical and sequential order. In the present study, government aided universities websites in Kolkata are included. The objective is to collect all the required data of the library websites during the period of December 2017 only. Link analysis as one of the webometric method is used for this study. The present study uses Google search engine for

collecting the required data. Beside this, lots of literatures are found about webometric study from where can get ideas about this study. The following search commands are used to collect data for each of the government aided universities websites in Kolkata as follows:

Table 1: Search Command and Retrieved Result in Google Search Engine

SL. No.	Search command	Results
1.	site:~abc	Total no of pages
2.	site: ~abc NOT link-domain: ~abc	Total No of external links or in-link/ R-WIF
3.	site: ~abc AND link-domain: ~abc	Total No of selflink
4.	site: ~abc filetype:pdf	Report total no of pdf file
5.	site: ~abc filetype:doc	Report total no of doc file
6.	site: ~abc filetype:ppt	Report total no of ppt file
7	site: ~abc filetype:xls	Report total no of xls file
8.	site:~abc (search in Google Scholar)	Total no of pages in google scholar

(Note: ~ denote Space among command)

(WP:Web page ; WIF:web impact factor ; IL: inlink; SL:self link)

The analysis of the data for the study was done by applying following statistical tools and techniques:

$$\text{R-WIF (Inlink-WIF)} = \frac{\text{Inlink (external backlinks) to the website}}{\text{Number of web pages published on the website which are indexed by a search engine}}$$

Spearman's Rank Correlation (R)

We used the Spearman's Rank Correlation (R) phenomenon in this study. Spearman's Rank Correlation (R) is used for measuring the strength of the relationship between different WIF (Simple WIF, Inlink WIF, Selflink WIF) of government-aided university websites in Kolkata. The formula used to measure the rank correlation is:

$$R = 1 - \frac{6\sum d^2}{n^3 - n}$$

d = Difference between the two ranks

n = Number of pairs of data (total observation)

Ranking Method

The ranking method has been used to arrange the university website as per their value of various web impact factors. Generally, give the smallest value the number 1, the next largest value the number 2, the next largest number 3 etc. The numbers 1, 2, 3... that are assigned to the various values are called the ranks.

WISER (Web indicators for science, technology and innovation research) indicator

The WISER indicator is disclosed by NAAC (National Assessment and Accreditation Council). To calculates WISER index value following formula is used:

The Formula of Retrieved WISER Value

WISER =log (visibility 50%) +log (size 20%) + log (Rich file15%) +log (scholer15%)

Table 2: WISER Value

WISER Value	
Visibility 50% (inlinks)	Size 20% (WebPages)
	Rich Files 15% (total pdf ,doc, xls,ppt)
	Scholars 15% (Google Scholar database)

Besides this, data calculation is done using MS Excel.

DATA ANALYSIS AND INTERPRETATION

Table 3 shows that WP, IL, SL, and sum of IL and SL which are retrieved through Google search engine using search Here calculate Inlink WIF (also called R-WIF) and rank according to Inlink WIF or R- WIF.

Table 3: Calculation of Inlink WIF/R-WIF (Web Impact Factor)

SL. No.	University	URL	WP	IL	SL	Inlink WIF/R-WIF	Rank
1.	Aliah University	www.aliah.ac.in	1380000	242000	95100	0.175362319	10
2.	University of Calcutta	www.caluniv.ac.in	168000	41200	47800	0.245238095	8
3.	Jadavpur University	www.jaduniv.edu.in	29300	5320	7040	0.181569966	9
4.	Netaji Shubhash Open University	www.wbnsou.ac.in	321000	54500	60800	0.169781931	11

SL. No.	University	URL	WP	IL	SL	Inlink WIF/R-WIF	Rank
5.	Presidency University	www.presiuniv.ac.in	10900	39800	62100	3.651376147	1
6.	Rabindra Bharati University	www.rbu.ac.in	1430000	1640000	173000	1.146853147	4
7.	Sanskrit College and University	www.sanskritcollegeanduniversity.org.in	1280	532	538	0.415625	6
8.	West Bengal National University of Juridical Science	www.nujs.edu	35700	80000	58100	2.240896359	2
9.	West Bengal University of Health Sciences	www.wbuhs.ac.in	159000	74600	82300	0.46918239	5
10.	West Bengal State University	www.wbsubregistration.org	62500	3040	3920	0.04864	13
11.	West Bengal University of Animal and Fishery Sciences	www.wbuafscl.ac.in	10300	688	846	0.066796117	12
12.	West Bengal University of Teachers, Training, Education Planning and Administration	www.wbuttepa.ac.in	5550	2050	2920	0.369369369	7
13.	Maulana Abul Kalam Azad University of Technology	www.wbut.ac.in	275000	382000	52100	1.389090909	3

Table 4, Present data of the total rich file with sum of PDF, DOC, XLS, and PPT. Beside this Google scholar database also used to collect data using predefined command.

Table 4: Rich File and Google Scholar Data

SL. No.	University	URL	WP	IL	SL	Total link	Rich Files (R)					Google scholar
							PDF	DOC	XLS	PPT	Total	
1.	Aliah University	www.aliah.ac.in	1380000	242000	95100	337100	16200	14	62	7	16552	14
2.	University of Calcutta	www.cal-univ.ac.in	168000	41200	47800	89000	35400	612	35	1	40116	612
3.	Jadavpur University	www.jaduniv.edu.in	29300	5320	7040	12360	566	40	6	1	582	40
4.	Netaji Shubhash Open University	www.wbnsou.ac.in	321000	54500	60800	115300	35700	25	59	9	35950	25
5.	Presidency University	www.presiuniv.ac.in	10900	39800	62100	101900	601	80	0	0	610	80
6.	Rabindra Bharati University	www.rbu.ac.in	1430000	1640000	173000	1813000	93300	13	192	149	95841	13
7.	Sanskrit College and University	www.sanskritcollegeanduniversity.org.in	1280	532	538	1070	195	0	0	0	195	0
8.	West Bengal National University of Juridical Science	www.nujs.edu	35700	80000	58100	138100	1890	46	10	1	1963	46
9.	West Bengal University of Health Sciences	www.wbuhs.ac.in	159000	74600	82300	156900	1380	6	8	0	1948	6
10.	West Bengal State University	www.wbsubregistration.org	62500	3040	3920	6960	1080	4	6	0	1125	4

SL. No.	University	URL	WP	IL	SL	Total link	Rich Files (R)					Google scholar
							PDF	DOC	XLS	PPT	Total	
11.	West Bengal University of Animal and Fishery Sciences	www.wbuafscl.ac.in	10300	688	846	1534	217	20	0	0	219	20
12.	West Bengal University of Teachers, Training, Education Planning and Administration	www.wbuttepa.ac.in	5550	2050	2920	4970	237	0	1	0	242	0
13.	West Bengal University of Technology	www.wbut.ac.in	275000	382000	52100	434100	154000	36	29	5	166734	36

Calculation of WISER index values of different university websites are mention below that is calculated by the revised indicator. Table 5: represent WISER Index value and ranking.

Table 5: Calculation of WISER Index Value and Ranking

SL. No.	University	URL	WISER INDEX VALUE	WISER Ranking
1.	Aliah University	www.aliah.ac.in	14.24	4
2.	University of Calcutta	www.caluniv.ac.in	14.58	3
3.	Jadavpur University	www.jaduniv.edu.in	9.91	9
4.	Netaji Shubhash Open University	www.wbnsou.ac.in	13.55	5
5.	Presidency University	www.presiuniv.ac.in	10.68	8
6.	Rabindra Bharati University	www.rbu.ac.in	15.82	1
7.	Sanskrit College and University	www.sanskritcollegeanduniversity.org.in	6.30	13
8.	West Bengal National University of Juridical Science	www.nujs.edu	11.76	6
9.	West Bengal University of Health Sciences	www.wbuhs.ac.in	11.49	7
10.	West Bengal State University	www.wbsubregistration.org	9.28	10
11.	West Bengal University of Animal and Fishery Sciences	www.wbuafscl.ac.in	7.84	11
12.	West Bengal University of Teachers, Training, Education Planning and Administration	www.wbuttepa.ac.in	7.62	12
13.	Maulana Abul Kalam Azad University of Technology	www.wbut.ac.in	15.15	2

Table 6: represent the calculation of rank correlation and using Spearman correlation and it is shown that WISER ranking and Inlink WIF ranking are not highly associated. whereas the value of r is 0.285714286.

Table 6: Correlation of WISER Ranking and Inlink WIF Ranking

SL. No.	University	URL	WISER Rank (x)	Inlink WIF Rank(y)	d=x-y	d^2
1.	Aliah University	www.aliah.ac.in	4	10	-6	36
2.	University of Calcutta	www.caluniv.ac.in	3	8	-5	25
3.	Jadavpur University	www.jaduniv.edu.in	9	9	0	0
4.	Netaji Shubhash Open University	www.wbnsou.ac.in	5	11	-6	36
5.	Presidency University	www.presiuniv.ac.in	8	1	7	49
6.	Rabindra Bharati University	www.rbu.ac.in	1	4	-3	9

SL. No.	University	URL	WISER Rank (x)	Inlink WIF Rank(y)	d=x-y	d^2
7.	Sanskrit College and University	www.sanskritcollegeanduniversity.org.in	13	6	7	49
8.	West Bengal National University of Juridical Science	www.nujs.edu	6	2	4	16
9.	West Bengal University of Health Sciences	www.wbuhs.ac.in	7	5	2	4
10.	West Bengal State University	www.wbsubregistration.org	10	13	-3	9
11.	West Bengal University of Animal and Fishery Sciences	www.wbuafscl.ac.in	11	12	-1	1
12.	West Bengal University of Teachers, Training, Education Planning and Administration	www.wbuttepa.ac.in	12	7	5	25
13.	West Bengal University of Technology	www.wbut.ac.in	2	3	-1	1
Total	260					

Below scatter diagram represent data association among WISER ranking and Inlink WIF ranking.

Correlation of WISER ranking and Inlink WIF ranking= 0.285714286

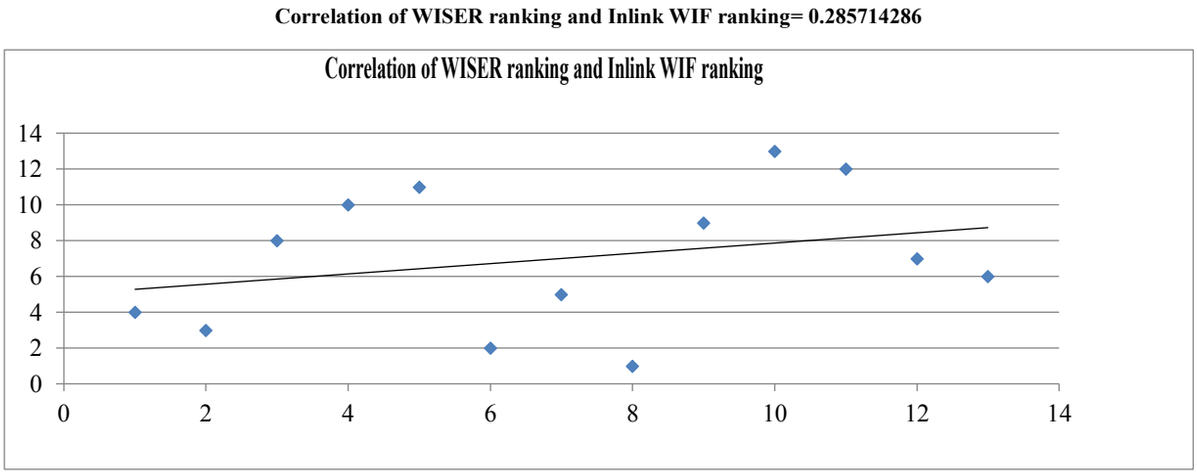

Fig. 1: Correlation of WISER Ranking and Inlink WIF Ranking

FINDING, CONCLUSION AND FURTHER RESEARCH

There are thirteen government aided universities in Kolkata included in this study. According to Inlink WIF Presidency University is ranked as a number one, whereas West Bengal State University ranked thirteen and WIF value of those Universities are 3.651376147 and 0.04864 respectively.

Correlation is a measure of the degree of association between two variables. For comparing two series of observation, It is sometimes necessary whether they are associated or not. It is denoted by the 'r' and its range is between +1 and -1; where +1 and -1 measure perfect position and negative association respectively r =0 implies no association.

Here correlation of WISER rank and Inlink WIF rank is 0.285714286, which present very low association among the WISER rank and Inlink rank.

A statistical process using a number that describes the degree of a linear relationship between two assets that either move together, or inversely, or are not related at all. The correlation coefficient is a way to measure the strength of the relationship between two assets, useful because analysis of one market can sometimes help us infer things about the other market. We use the correlation phenomenon in our analyses and alerts.

This study has been searching and there is an extent for prospect webometric investigation in this area. Web presence and links analysis countries are gifted in general terms, but in practice, web presence and links are not entirely equivalent to citations in the scholarly as much of the web content of country.

Further research may be possible using the more statistical technique. For further research, regression formula may be used to calculate dependency of different link data. The

Multiple regressions may be used to predict web link that is dependent on each other. The standard error may be used to estimate how far the sample mean of a different web link is likely to be from the population mean. After that ANOVA may be used to test the difference between different Web Impact factors.

REFERENCES

Agusten, D. S. T., & Winarti, S. (2013). Analysis on backlinks and pagerank of automotive company websites using crawler tool. *World of Computer Science and Information Technology Journal (WCSIT), 3*. Retrieved from https://pdfs.semanticscholar.org/8f47/0b0f8bfa2db8f1f2c70c3ff49b4ddf056d5c.pdf.

Ahuja, M. S., Bal, J. S., & Varnika. (n.d.). Web crawler: Extracting the web data. *International Journal of Computer Trends and Technology (IJCTT), 13*(3). Retrieved from http://www.ijcttjournal.org/Volume13/number-3/IJCTT-V13P128.pdf

Andersson, E., & Ekstrom, P. (2004). Investigating Google's Pagerank Algorithm (Rep.). Retrieved http://www.it.uu.se/edu/course/homepage/projektTDB/vt04/projekt5/website/report.pdf

Arshad, A., & Ameen, K. (2015). Usage Patterns of Punjab University Library Website: A Transactional Log Analysis. *Emeraldinsight, 33*. Retrieved from http://www.emeraldinsight.com/doi/full/10.1108/EL-12-2012-0161

Bershadskaya, M., Voznesenskaya, Y., & Karpenko, O. (2016). Webometrics ranking in the context of accessibility of higher education. *Universal Journal of Educational Research, 4*(7). doi:10.13189/ujer.2016.040702

Brin, S., & Page, L. (n.d.). The Anatomy of a Large-Scale Hypertextual Web Search Engine. Retrieved from http://ilpubs.stanford.edu:8090/361

Das, N. G. (2013). *Statistical methods* (9th ed.). New Delhi: Mc Graw Hill Education (India) Private limited.

Didegah, F., & Erfanmanesh, M. A. (n.d.). The study of malaysian public universites' performance on the world web. *Emerald Insight, 27*. Retrieved from http://www.emeraldinsight.com/doi/pdfplus/10.1108/07419051011070635

Google Search. (2018). Retrieved April 17, 2018, from https://en.wikipedia.org/wiki/Google_Search

Jalal, S. K., Sutradhar, B., & Mukhopadhyay, P. (2016). Webometrics analysis of top ten Asian and Indian universities. *SRELS Journal of Information Management, 53*(2). doi:10.17821/srels/2016/v53i2/91271

Janbandhu, R., Dahiwale, P., & Raghuwanshi, M. (n.d.). Analysis of web crawling algorithms. *International Journal on Recent and Innovation Trends in Computing and Communication, 2*. Retrieved from https://www.researchgate.net/publication/268357163_Analysis_of_web_crawling_algorithms.

Jayshankar, R., & Sujitha, M. (n.d.). *Webometrics tools and techniques: An overview*.

Khan, A., & Idrees, H. (2015). *Calculating web impact factor for university websites of Pakistan.* Emeraldinsight, *33*. Retrieved from http://www.emeraldinsight.com/doi/pdfplus/10.1108/EL-01-2014-0022

Leydesdorff, L., & Curran, M. (n.d.). Mapping university-industry-government relations on the internet: The construction of indicators for a knowledge-based economy. *International Journal of Scientometrics, Informetrics and Bibliometrics, 4*. Retrieved from https://www.researchgate.net/publication/28090935_Mapping_University-Industry-Government_Relations_on_the_Internet_the_construction_of_indicators_for_a_Knowledge-Based_Economy.

Madhusudhan, M., & Prakash, S. (2013). Websites of Indian institutes of technology: A webometric study. *International Journal of Library and Information Studies, 3*(4). Retrieved from http://www.ijodls.in/uploads/3/6/0/3/3603729/55-65313.pdf

Pranav, A., & Chauhan, S. (n.d.). Efficient Focused Web Crawling Approach for Scearch Engine. *International Journal of Computer Science and Mobile Computing(IJCSMC), 4*. Retrieved from http://www.ijcsmc.com/docs/papers/May2015/V4I5201599a17.pdf

Shafi, S. M., & Bhat, M. H. (2014). Performance and visibility of indian research institutions on the web. *Emeraldinsight, 44*. Retrieved from http://www.emeraldinsight.com/doi/full/10.1108/VINE-06-2013-0034

Thelwell, M. (2011). *A Comparison of Link and URL Citation Counting, 63*. Retrieved from http://www.emeraldinsight.com/doi/pdfplus/10.1108/00012531111148985

Wills, R. S. (2006). Google's Pagerank the Math Behind the Search Engine. Retrieved from https://9b90bb5f-a-62cb3a1a-s-sites.googlegroups.com/site/rebeccawillswebsite/Home/papers_and_presentations/rsw_06_google_intelligencer.pdf?attachauth=ANoY7cqVdB7BiKtVpH9P-aMHsPAxq_dIpfG-55GHY8XQWV4S3fOpTeZfyvIRyg0WAsT7VAY-SX5gL_RG9SiAwf_CBhQUbhdC-9Rcotg41R40XxjNpH4cnH_6wxLCESLMw8yYR82AV5M2JKSg7jrx-P7tmPeFCA5eBc34LarRn9jAb_qM14zRfXjUs_rvC4fDb3sl_oKW3OBFDtHjgexBYjQOLHKIM-6HHF2t5Xxky11wFpCHyQYMMn53wRGwaNx-dxbkESPsOx71glslNeNO4oJKR6T6Qsy_zWCaVRDQHTK_8ZYKhsNV18PtOmk%3D&attredirects=0.