Proceedings of the
13th International Conference on Cyber Warfare and Security

National Defense Univeristy
Washington DC, USA
8-9 March 2018

Edited by
Dr. Jim Q. Chen and Dr. John S. Hurley

acpi

A conference managed by ACPI, UK

Proceedings of the

13th International Conference on Cyber Warfare and Security

ICCWS 2018

Hosted By

National Defense University

Washington DC, USA

8 - 9 March 2018

Edited by

Dr John S. Hurley and Dr Jim Q. Chen

Review Process
Papers submitted to this conference have been double-blind peer reviewed before final acceptance to the conference. Initially, abstracts were reviewed for relevance and accessibility and successful authors were invited to submit full papers. Many thanks to the reviewers who helped ensure the quality of all the submissions.

Ethics and Publication Malpractice Policy
ACPIL adheres to a strict ethics and publication malpractice policy for all publications – details of which can be found here: http://www.academic-conferences.org/policies/ethics-policy-for-publishing-in-the-conference-proceedings-of-academic-conferences-and-publishing-international-limited/

Conference Proceedings
The Conference Proceedings is a book published with an ISBN and ISSN. The proceedings have been submitted to a number of accreditation, citation and indexing bodies including Thomson ISI Web of Science and Elsevier Scopus.

Author affiliation details in these proceedings have been reproduced as supplied by the authors themselves.

The Electronic version of the Conference Proceedings is available to download from DROPBOX https://tinyurl.com/ICCWS2018 Select Download and then Direct Download to access the Pdf file. Free download is available for conference participants for a period of 2 weeks after the conference.
The Conference Proceedings for this year and previous years can be purchased from http://academic-bookshop.com

E-Book ISBN: 978-1-911218-73-9
E-Book ISSN: 2048-9889
Book version ISBN: 978-1-911218-74-6
Book Version ISSN: 2048-9870

Published by Academic Conferences and Publishing International Limited
Reading
UK
44-118-972-4148
www.academic-publishing.org

Nostradamus Ratios: Why is Russia an Outlier?

Michael Bennett Hotchkiss
Independent Researcher, Preston CT, USA

Abstract: Webometrics is defined as *"the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the web drawing on bibliometric and informetric approaches"* (Bjorneborn & Ingwerson 2004); and *"the study of web-based content with primarily quantitative methods for social science research goals"* (Thelwall 2009). Hit Count Estimates (HCEs) are a webometric provided by search engines giving an approximate number of the relevant pages indexed by the search engine. HCEs are not very reliable, but they have been observed to be up to 78% reliable on Google and it is the recommended platform for HCE-based research (Uyar 2009). Webometrics may have promise in the fight against online disinformation and influence campaigns as evidenced by tools like Hamilton 68; or in the application of Google Trends and similar tools to highlight the possible effects of "information attacks" (Hotchkiss 2017). The current investigation seeks to apply HCEs as a webometric which may supplement existing findings in disinformation research. HCEs were collected over a period of 21 days for 38 "Western" and Russian news sites and 38 'Top Level Domains' (TLDs) in order to create averages which would account for observed daily variability in HCEs. HCE results were collected for the total number of pages on the news site or TLD, the number of times the search term "Nostradamus" (or regional translation) appeared on pages on the news site or TLD; and the number of times "Nostradamus" (or regional translation) appeared in the page title on news sites. These data points were used to create various simple ratios and rankings. In addition, the qualitative aspects of the news sites with Nostradamus in page titles were analysed to determine if a site promoted belief in the supernatural powers of Nostradamus. Findings suggest Russia is an outlier among tested news sites and TLDs, and it promotes Nostradamus as a genuine prophet (especially to Spanish-language audiences via RT.com). Ultimately however, the sometimes apparently chaotic and non-contextual nature of HCEs means they are probably only good for supplemental findings, at best, and should be accompanied by a qualitative analysis if possible.

Keywords: disinformation, fake news, Google trends, Hamilton 68, Hit Count Estimates, webometrics

1. Introduction

Search engines (such as Bing and Google), return approximate numbers of results for a query, known as 'hit count estimates' (HCEs). Prior research has shown that these estimates can be up to 78% accurate on Google for queries using a single term which return under 1000 results, though that accuracy plummets as additional terms are added to the query, and it cannot be definitively determined if that accuracy is generalizable to larger queries since it cannot be practically tested, although larger HCEs seem to have similar error characteristics to smaller ones (Uyar 2009). Webometrics researchers have made consistent findings that Google produces the most reliable HCEs and therefore it is the recommended platform for investigations regarding them in the social sciences (Thelwall 2008, Uyar 2009).

For this paper, HCEs were collected over a period of 21 days between July 23 and August 27, 2017 from 38 different online news sites and 38 top level domains (TLDs) on Google for the search term "Nostradamus". The sites selected were intended to represent a mix of Western and Russian journalism sites, as well as a broad geographic sample of regions in order to compare the relative proportion of the appearance of the term. (HCEs for 'in title' instances of the search term were only collected for the news sites.) In addition, the content of the news sites was qualitatively examined to determine whether they classified Nostradamus as worthy of credibility or scepticism.

The goal was to determine if there were quantitative differences between the proportion of these results at different sites and if such hypothetical differences might provide some kind of insight into intent once examined qualitatively. It was the author's hypothesis based on prior research that Russia would be associated with increased promotion of Nostradamus as a concept in the popular consciousness.

2. Literature review

Webometrics, a facet of informetrics (the study of quantitative aspects of information), has been defined as: "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches." (Bjorneborn & Ingwerson, 2004). Thelwall (2009) defined webometrics as "the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study". Webometrics may

include research from beyond information sciences including communication studies, statistical physics, and computer science (Thelwall, Vaughan, & Bjorneborn, 2006).

By these definitions, Google Trends (and its predecessor, Google Zeitgeist) is a 'webometric' tool because it indexes and applies statistical quantification to web data (in this case user search queries) in a way that can be used to support decision making. Google Trends has received considerable attention because of its perceived utility as a portal into the collective consciousness which can be of use in marketing, investing, psychology, and epidemiological research (see Google Flu Trends and successor tools like AutoRegression with GOogle search data (ARGO)).

In combination with contextual blog searches, Thelwall (2009) showed how as a case study in webometrics, Google Trends could be used to track the "Danish cartoon" (crisis) of 2005. This has been questioned as a possible Russian "active measure" in the past by credible Kremlinologists (Boghardt 2006). The Russian angle is not something explored by Thelwall, but his example is useful and illustrative for this purpose.

The potential of formalized approaches to using informetrics to detect, expose, and discredit disinformation is already being realized through intelligence-oriented webometric tools like 'Hamilton 68' by the German Marshall Fund. Introduced in summer 2017, the dashboard tool uses a proprietary method to track accounts associated with Russian information operations on Twitter in real time, and promises to identify disinformation campaigns as soon as they begin (Rosenberger & Berger 2017).

Other examples of potential applications of webometrics which have come into focus during the course of public investigations into so-called "Russian election interference" includes studies into the number of users who can be targeted with precision using Facebook advertising tools. In short, there is ample evidence that webometrics is an emerging area of complementary research for investigators who are seeking to expose internet-based influence and disinformation schemes.

Prior research by the author has argued that the 2001 'search anomaly' which saw "Nostradamus" become an unlikely top 10 search term on 11 September 2001 on Google and the #1 search term of the year on Google, Lycos, and Yahoo is the likely result of Russian 'active measures' and cyber operations. Similar patterns have been observed in Google Trends data from Poland (2010), Ukraine (2014), and Hungary (2015) during provable periods of Russian geopolitical involvement in those countries' affairs (Hotchkiss 2017a).

Beyond this, the background of the 1999 Russian terror attacks may honour elements of Russian state mythology when viewed through the lens of Erika Cheetham's controversial interpretation of Nostradamus' Century X Quatrain 72 prophecy about 1999 (Hotchkiss 2017b). The use of the prophecies of the 16th century 'seer' Nostradamus for psychological warfare and disinformation were widespread in European history, including during World War II (Wilson 2007). Before Perestroika, references giving Nostradamus credibility in terms of his predictive ability for "revolutions" and "political assassinations" found their way into a Soviet journal which would have been state-censored (JPRS 1978). Christensen (1998) found in a 1994 survey of Ukrainians and Belarussians that these populations ascribed major explanatory significance for 'apocalyptic' events to the prophecies of Nostradamus (such as the Chernobyl disaster), in ways which were disproportionate with the West. This was related to broader patterns of Russian apocalyptic thinking and beliefs (similar to those espoused by Alexander Dugin today). This combination of webometric data and historical research provides strong evidence to suggest Russia's involvement in the covert promotion of Nostradamus prophecies in the context of modern cyber and information warfare.

The current research seeks to assess whether Google Search hit count estimates (HCEs) can be used for similar purposes as a supplementary webometric tool that may be able to expose information operations (albeit not in real-time). HCEs are a webometric that often appear on results pages, and are intended to give the user an idea of the approximate total results for a given query. Since the late 1990's, researchers have been interested in the accuracy of these counts and the decisions, if any which can be made from them. While researchers consistently found that these are not exact counts, and should be approached cautiously for decision-making, prior research from2008 and 2009 has shown that Google produced the most accurate HCEs (Thelwall 2008, Uyar 2009).

For queries returning fewer than 1000 results, Google was approximately 78% accurate when a single term was used, but as additional terms were added, the accuracy of the results decreased substantially. Furthermore,

because of the difficulty of verifying such results for larger samples of data, and because search engines typically only return 1000 results viewable to the user, it is unknown if these findings about accuracy in HCEs are generalizable to samples of greater than 1000 results (Uyar 2009).

3. Method

3.1 Investigation A: Quantitative and investigation B: Quantitative

Over 21 days between 23 July 2017 and 27 August 2017 between 12 PM EST and 10 PM EST, Google HCEs were collected using Google Chrome on the desktop for 38 news portals: 'Investigation A' (Table 1) and 38 TLDs 'Investigation B' (Table 3). (These findings were recorded in Excel and all subsequent calculations were performed using Excel.)

For the first investigation involving news sites, sites were selected to include a mix of state news agencies, popular mainstream news sites, and news sites known to be tied to Russia. The goal was to have a sample of over 30 total sites and no site which was considered was excluded.

Results were collected for the commands:

- site:newssite.com,
- site:newssite.com "Nostradamus", and
- site:newssite.com intitle:"Nostradamus"

For the second investigation involving TLDs, results were collected for:

- site:*.tld, and
- site:*.tld "Nostradamus"

Because TLDs roughly correspond to geographic locations, a selection was chosen which corresponded to a broad swathe of Europe, Asia, and the Americas in order to provide a global picture. (It was observed that the query to produce "intitle" results for the TLD was immediately unreliable and so it was decided to abandon collection of those results.)

In both investigations, if the source did not primarily use the Latin alphabet, the equivalent translated term in the regional alphabet for Nostradamus was used (regional spellings in non-ASCII languages such as Cyrillic:"Нострадамус" or Chinese: "諾查丹瑪斯"). (The exception in this case are both Koreas; South Korea includes considerable results in the Latin alphabet but very few in the Korean. North Korea only returns one result on Nostradamus and is in the French language, but it returns none in Korean. Moldova (*.md) also produces considerably more results when searched using the Latin vs Cyrillic alphabet, but in this case Cyrillic was used for consistency.)

Quotes were used around "Nostradamus" (or equivalent translation) in an effort to return the exact term and avoid semantic matching.

HCEs can change page to page on results, thus to keep collection standardized, the HCE was taken from page 1 of results.

In order to minimize error due to day to day variation in reported results, results were collected over a period of 21 days and then averaged. The use of a timeseries to create a more accruate average using HCEs over a period of time has been an approach since the late 1990s (Rousseau 1999).

Different measures which could be used for ranking based on the collected data were considered to be:

- The average HCE of the term on the site or TLD. (Investigation A and Investigation B)
- *This may be an indicator of the overall interest in the term on the site relative to other sites, in terms of who has the "most" content related to the term. It could also be an indicator of the "user culture" on the site or TLD.*
- The average HCE of the term in titles of pages on the site or TLD. (Investigation A)

- *This may be an indicator that the site editors intended to publish content related to the term rather than content which was created by site users.*
- The ratio of the average HCE of the term on the site or TLD to the average number of pages returned as a HCE for the site or TLD. (Investigation A and Investigation B)
- *This gives an idea of the 'average term per page' on the site. This may be useful for making comparisons about the proportion of content which is dedicated to the term on a given site or TLD. It also takes into account the number of pages on the sites/TLDs so that those with less total pages could theoretically stand out in the ranking based on the proportion of content.*
- The ratio of the average HCE of the term in titles of pages on the site or TLD to the average number of pages returned as a HCE for the site or TLD. (Investigation A)
- *Similar to the 'average term per page' metric, and with the assumption that intent of the site editor is less ambiguously when it comes to page titles than general site content, the 'average term intitle' metric may be a more accurate metric in determining 'intention' behind the use of the term as content on the site.*
- A 'superranking' ranking based on averaging the applicable collected ranks. This rank would ideally factor as many of the ranks as possible into a single rank, since each of the ranks may measure different aspects of magnitude related to the term.

3.2 Investigation A: Qualitative (Column B in Table A)

It would be impractical – if not impossible (because of 1000 page result return limits) to examine every site reference of Nostradamus on each domain as returned in search results, and in addition there are many more confounding complexities when not searching 'intitle' for results. It is also simply common-sense to assume that when a subject is included in the title of a page, that it is linked to the content of the page. Sometimes, a user for example might write "Nostradamus" in comments which becomes indexed in the engine (Zerohedge?). For these matters of practicality, and common-sense error reduction, only pages which had an "intitle" mention of Nostradamus were examined qualitatively.

The intitle search results for the news sites from Investigation A were examined qualitatively to determine if they (even once) overtly promoted Nostradamus in a supernatural sense. Those remaining were classified as to whether they are sceptical of Nostradamus, neither promoted nor were sceptical of Nostradamus (neutral), promoted plus were sceptical of Nostradamus (mixture), or had no content about Nostradamus (no content).

Google Translate was used to read non-English pages.

3.3 Data curation

The author elected to remove the 27 August 2017 day from all samples as it included an extraordinary amount of clear outliers. However, all other data, including presumed outlier "errors" in the HCEs for the other days were maintained, with the hope that errors would be "averaged out" by the remaining 20 day sample. The removal of August 27 did not seem to alter overall rankings.

Removal of outliers was tested to see if it improved the overall correlation in the ranking measures.

4. Results

The full data file can be accessed at: https://drive.google.com/open?id=0BzEHLGuzsQA9WmFJU1pkVHd4Wk0

5. Discussion

5.1 Prominent oddities

In the course of collecting this data, and becoming acquainted with the typical results, it was noted that some of the news sites in particular experienced large fluctuations in the consistency of HCEs. This was true for example in the case of Vice.com which would jump from ~1 million to ~10 million total page HCEs returned on a given day (5,090,900 (mean) 8,760,000 (mode) 5,050,000 (median) 904,000 (min) 10,400,000 (max), and 9,496,000 (range)). USAToday.com also experienced similarly strange fluctuations, as did Xinhuanet.com (Chinese), though neither fluctuated as reliably as Vice.com.

Table 1: Investigation A - quantitative and qualitative (sorted by 'superrank')

News Site	Qualitative (Intitle)	Average Number of Pages	Term Avg. (A)	Rank (A)	Intitle Avg. (B)	Rank (B)	Term Avg. Ratio (C)	Rank (C)	Intitle Avg. Ratio (D)	Rank (D)	Super-rank
Pravdareport.com	Promoted	33,495	1384.5	5	27.9	4	0.0413345275	1	0.0008329601	2	**1**
LeMonde.fr	Neutral	919,150	2095	3	50.3	2	0.0022792798	7	0.0000547245	3	**2**
Pravda.ru *(Cyrillic)	Promoted	335,300	2706.5	1	17.65	9	0.0080718759	3	0.0000526394	4	**3**
RT.com	Promoted	605,700	2101.5	2	18.2	8	0.0034695394	5	0.0000300479	5	**4**
English.pravda.ru	Promoted	13,650	78.05	26	21.75	6	0.0057179487	4	0.0015934066	1	**5**
Newsweek.com	Neutral	133,700	1431.45	4	2	25	0.0107064323	2	0.0000149589	7	**6**
theguardian.com	Mixture	2,786,500	513.05	7	16.1	10	0.0001841199	11	0.0000057779	14	**7**
LATimes.com	Mixture	5,685,000	546.8	6	28.95	3	0.0000961829	20	0.0000050923	15	**8**
Vice.com	No content	5,090,900	380.2	12	76.1	1	0.0000746823	24	0.0000149482	8	**9**
CNN.com	Sceptical	3,477,500	511.6	8	13.75	13	0.0001471172	14	0.0000039540	16	**10**
Newsweek.pl	Promoted	297,450	149.25	21	5	16	0.0005017650	8	0.0000168095	6	**10**
Huffingtonpost.com	Promoted	1,750,000	256.2	18	14.3	12	0.0001464000	15	0.0000081714	11	**12**
SMH.com.au	Promoted	1,624,500	282	17	9.95	14	0.0001735919	12	0.0000061250	13	**12**
TheSun.co.uk	Promoted	635,000	135.25	22	5	16	0.0002129921	10	0.0000078740	12	**14**
ibtimes.com	Promoted	352,300	170.15	20	2.9	23	0.0004829691	9	0.0000082316	10	**15**
NYTimes.com	Sceptical	8,510,500	434.95	10	22.2	5	0.0000511075	29	0.0000026085	20	**16**
BBC.co.uk	Neutral	9,386,000	481.45	9	20.2	7	0.0000512945	28	0.0000021521	23	**17**
USAToday.com	Sceptical	5,351,000	286	16	15.7	11	0.0000534480	27	0.0000029340	19	**18**
Time.com	Sceptical	761,200	101	23	3	20	0.0001326852	17	0.0000039411	17	**19**
ZeroHedge.com	No content	152,200	414.15	11	0	31	0.0027210907	6	0.0000000000	31	**20**
telegraph.co.uk	Mixture	4,836,500	289.55	15	8.85	15	0.0000598677	26	0.0000018298	24	**21**
Washingtonpost.com	Sceptical	4,183,500	358.25	13	4.45	18	0.0000856340	22	0.0000010637	27	**21**
DailyMail.co.uk	Neutral	3,439,000	311.95	14	3	20	0.0000907095	21	0.0000008723	28	**23**
Izvestia.ru *(Cyrillic)	Neutral	210,560	26.6	32	2.15	24	0.0001263298	18	0.0000102109	9	**23**
Foxnews.com	Promoted	576,500	96.1	24	0.95	29	0.0001666956	13	0.0000016479	26	**25**
DW.com	Promoted	1,197,500	76.8	27	3	20	0.0000641336	25	0.0000025052	21	**26**
NBCNews.com	Sceptical	506,850	68.9	28	1.1	28	0.0001359377	16	0.0000021703	22	**27**
ABCNews.go.com	Neutral	577,400	45.5	29	1.95	27	0.0000788015	23	0.0000033772	18	**28**
Reuters.com	Neutral	11,140,000	199.35	19	3.8	19	0.0000178950	33	0.0000003411	29	**29**
forbes.com	No content	793,350	91.75	25	0	31	0.0001156488	19	0.0000000000	31	**30**
WSJ.com	Sceptical	1,095,850	40.4	30	2	25	0.0000368664	30	0.0000018251	25	**31**
Xinhuanet.com *(Chinese)	No content	1,222,200	39.1	31	0.05	30	0.0000319915	31	0.0000000409	30	**32**
CBC.ca	No content	954,950	26.55	33	0	31	0.0000278025	32	0.0000000000	31	**33**
TASS.ru *(Cyrillic)	No content	381,600	1.7	34	0	31	0.0000044549	34	0.0000000000	31	**34**
TASS.com	No content	59,730	0	35	0	31	0.0000000000	35	0.0000000000	31	**35**
thenews.pl	No content	94,050	0	35	0	31	0.0000000000	35	0.0000000000	31	**35**
unian.info	No content	32,425	0	35	0	31	0.0000000000	35	0.0000000000	31	**35**
Xinhuanet.com/English	No content	326,350	0	35	0	31	0.0000000000	35	0.0000000000	31	**35**

Table 2: Investigation A – correlation coefficients (no outliers removed)

Term vs. intitle (pre-ratios)	r = 0.438270565	**Term vs. Intitle ratios**	r = 0.52647567
Pearson (raw data A, B)	*n = 38*	*Pearson (raw data C, D)*	*n = 38*
Term vs. Intitle (pre-ratios)	r = 0.754718562	**Term vs. Intitle ratios**	r = 0.778469774
Spearman's rho (ranks A, B)	*n = 38*	*Spearman's rho (Ranks C, D)*	*n = 38*

Table 3: Investigation B - quantitative (sorted by 'superrank')

Top Level Domain (TLD)	Country	Average Number of Pages	Term Avg. (A)	Rank (A)	Term Avg. Ratio (B)	Rank (B)	Superrank
*.ru (Cyrillic)	Russia	502,850,000	711,650	2	0.001415233	4	**1**
*.cn (Chinese)	China	351,150,000	469,750	3	0.001337747	5	**2**
*.ro	Romania	81,200,000	153,170	7	0.00188633	3	**3**
*.ua (Cyrillic)	Ukraine	78,905,000	104,915	10	0.001329637	7	**4**
*.jp (Japanese)	Japan	716,200,000	361,200	4	0.000504328	14	**5**
*.ir (Persian)	Iran	33,985,000	83,595	16	0.002459762	2	**5**
*.kr	South Korea	133,200,000	111,100	9	0.000834084	10	**7**
*.mk (Cyrillic)	Macedonia	17,100,000	64,160	19	0.003752047	1	**8**
*.sk	Slovakia	89,750,000	83,995	15	0.000935877	9	**9**
*.mx	Mexico	109,385,000	79,230	17	0.000724322	11	**10**
*.id	Indonesia	61,340,000	61,025	21	0.000994865	8	**11**
*.dk	Armenia	227,150,000	88,500	14	0.00038961	17	**12**
*.hu	Hungary	120,150,000	66,110	18	0.000550229	13	**12**
*.it	Italy	909,100,000	144,050	8	0.000158453	24	**14**
*.com	N/A	25,270,000,000	2,840,500	1	0.000112406	32	**15**
*.fr	France	1,255,000,000	168,250	6	0.000134064	27	**15**
*.pl	Poland	400,100,000	102,285	12	0.000255649	21	**15**
*.ar	Argentina	97,485,000	55,910	23	0.000573524	12	**18**
*.de	Germany	1,966,500,000	227,150	5	0.00011551	31	**19**
*.si	Slovenia	290,050,000	62,855	20	0.000216704	22	**20**
*.cz	Czech Republic	197,000,000	54,925	24	0.000278807	18	**20**
*.cu	Cuba	1,969,500	2,622	36	0.001331302	6	**20**
*.au	Australia	776,950,000	94,730	13	0.000121925	30	**23**
*.uk	United Kingdom	1,704,500,000	104,495	11	6.13054E-05	35	**24**
*.za	South Africa	166,100,000	42,860	27	0.000258037	20	**25**
*.ve	Venezuela	28,905,000	13,373	32	0.000462636	15	**25**
*.ee	Estonia	78,095,000	21,735	29	0.000278315	19	**27**
*.ge (Georgian)	Georgia	27,850,000	11,705	33	0.000420287	16	**28**
*.in	India	358,250,000	50,285	25	0.000140363	25	**29**
*.nl	Netherlands	477,800,000	60,095	22	0.000125774	29	**30**
*.fi	Finland	128,110,000	20,745	30	0.000161931	23	**31**
*.my	Malaysia	125,545,000	17,460	31	0.000139074	26	**32**
*.ie	Israel	472,500,000	48,696	26	0.00010306	33	**33**
*.md (Cyrillic)	Moldova	22,410,000	2,990	35	0.000133423	28	**34**
*.ca	Canada	1,064,000,000	40,245	28	3.78242E-05	36	**35**
*.by (Cyrillic)	Belarus	63,985,000	5,680	34	8.87708E-05	34	**36**
*.ph	Philippines	79,380,426	2,319	37	2.92075E-05	37	**37**
*.kp	North Korea	36,900	1	38	2.71003E-05	38	**38**

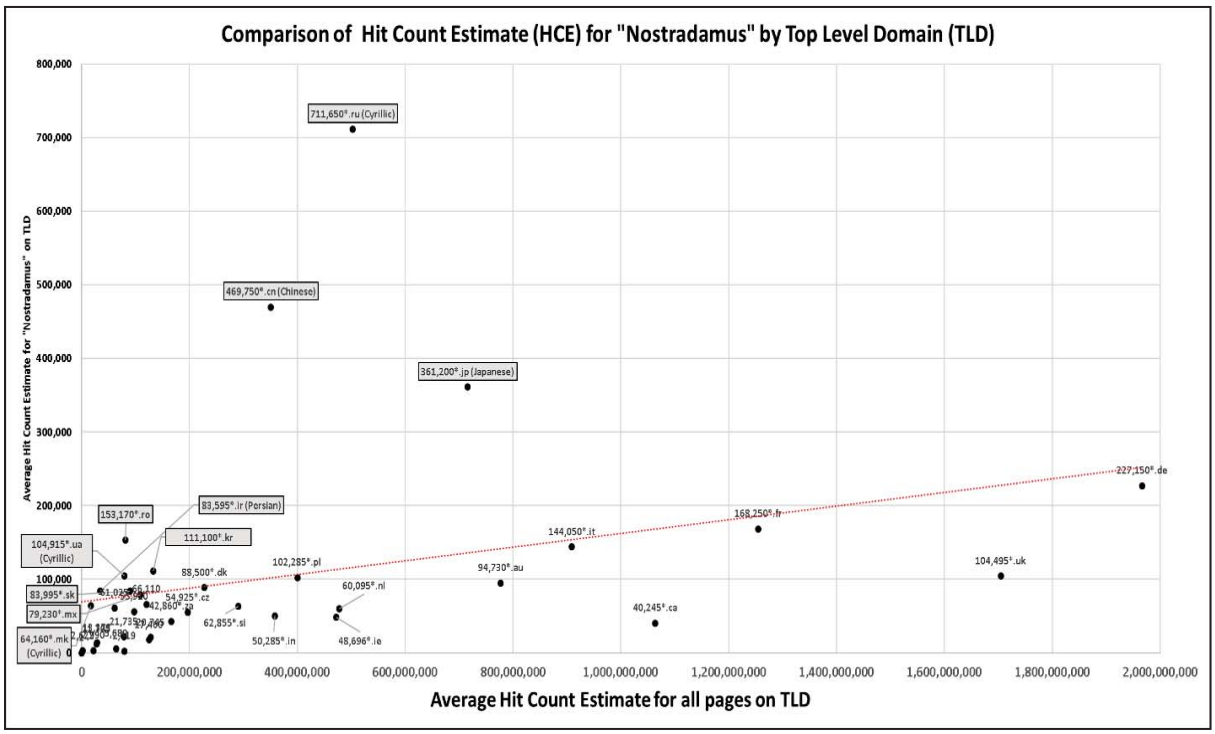

Figure 1: Investigation B - Average HCE for all pages on TLD plotted against average HCE for "Nostradamus" mentions on pages (includes regional translations; Top 10 superranked TLDs highlighted and *.com TLD omitted). This view highlights Russia as a clear outlier in the analysis

Table 4: Investigation B – correlation coefficients

Average Number of Pages vs. Term Avg. - All results	
Pearson Correlation	r = 0.959065
Spearman's rho	r = 0.627312
	n = 38
Average Number of Pages vs. Term Avg. - Excludes lowest superrank and top 5 superranks	
Pearson Correlation	r = 0.997066
Spearman's rho	r = 0.690845
	n = 31

Throughout the process of collecting results, RT.com experienced a consistent decline in the number of on-page references to Nostradamus (from over 4000 results estimated on the day prior to officially collecting results, to 1300 on August 23 (this decline did not materially change RT's ranking)). Conversely, despite not having pro-Nostradamus content, Newsweek's number of results grew from 877 at the start of collection to as high as 2300 results on August 16.

By comparison, the number of average "intitle" results was relatively stable throughout the process (though Ibtimes.com seemed to experience a consistent decline from 4 to 2 intitle results). Total page estimates were second most consistent, and on page mentions of Nostradamus were least consistent.

5.2 Findings investigation A: Quantitative

Russian news sites seem to mention "Nostradamus" more frequently than Western sources based on Table A. Russian sites hold 4 of the top 5 superrank positions. (The only other site in the top 5 that is not Russian (Le Monde) does not have qualitative results which support beliefs in Nostradamus (usually it is history and recipes), while the Russian sites do (often) support beliefs in Nostradamus). (Table 1)

Without removal of outliers, there is a significant Spearman's Rho of moderate strength for the relationships between the ranks of raw HCEs for 'term on page' (A) and 'term intitle' (B) (r = 0.754718562, n = 38) and HCE ratios for 'average term per page' (C) and 'average term intitle' (D) (r = 0.778469774, n = 38). (Table 2)

It should be noted that English.pravda.ru seems to have been rolled into Pravdareport.com but they produce different results when queried by Google.

5.3 Findings investigation A: Qualitative

Examining the actual stories on each of the sites which include "Nostradamus" (or "Нострадамус" for the *.ru sites) provides additional context. These results give some insight which affirm some quantitative results, but also dispute others from a contextual perspective.

For example, Vice.com showed an average of 76.1 pages with Nostradamus in the title of the page (the highest number of the collected sample) throughout the data gathering phase. However, when the pages were examined for context, it was found that there were no pages on the entire site with such a title reference which actually discussed Nostradamus. All of the pages were subject landing pages, or possibly user pages, but no content about Nostradamus was included. Vice.com would appear much lower in the rankings and should possibly be removed from this analysis based on these findings.

Similarly, CNN included on the average 13.75 in title results for Nostradamus. All but one of the pages is actually written by CNN officially. The one official page is definitely sceptical of Nostradamus, but the remaining pages returned are "iReport" pages which can be blogged by CNN users. They have value similar to Wikipedia from a content quality perspective and do not represent the editorial opinion of CNN. Thus, CNN does host pro-Nostradamus pages, but those pages are not reflective apparently of the CNN official editorial opinion. CNN was therefore classified as sceptical of Nostradamus, despite hosting pro-Nostradamus content on the majority of the pages which were returned in search results. CNN.com would appear lower in the rankings based on these findings if the iReport pages were not considered.

Other observed findings which were unique to some sites include duplicate pages on multiple subdomains of the same news site which overall would 'amplify' the results for those sites. (Newsweek.pl for example.)

Sites which were found to most likely promote Nostradamus as "real" include: DW.com, English.pravda.ru, Foxnews.com, Huffingtonpost.com, Ibtimes.com, Newsweek.pl, Pravda.ru, Pravdareport.com, RT.com, and SMH.com.au. (With the exception of one RT story in English about an "Italian Nostradamus", the stories on RT.com and DW.com which treat Nostradamus with credulity are in Spanish.)

LATimes.com was classified as a "mixture" because while all of the English-language content was sceptical of Nostradamus (including LA-region earthquake predictions), it also hosts at least 2 Spanish-language stories which treat Nostradamus with credibility – including specific claims which say he predicted earthquakes. Thus, English and Spanish readers of the site may come away with very different experiences. (TheGuardian.com and Telegraph.co.uk were classified as "mixture" because they contain books which promote Nostradamus in their online bookstores, though the site content does not advocate for his supernatural abilities.)

Foxnews.com is listed as a "promoter" of Nostradamus because the one page it hosted was just a page title with no content, but 'TheHill.com' story which the link referred to ("Nostradamus' of Middle East predicts unprecedented crisis for Obama in 2011") treats the concept of Nostradamus (and the subject Middle Eastern prophet with credulity rather than scepticism). Despite the finding of "promoter", this is overall a weak finding. Similarly SMH.com.au has one page which promotes a belief in Nostradamus from 2005. While most content is neutral on the site, it was classified as a promoter for having a single instance of pro-Nostradamus content.

5.4 Findings investigation B: Quantitative

At the TLD level, there is a strong relationship between the number of pages on a TLD and the number of times it mentions Nostradamus (or the regional spelling). Russia seems to have the most mentions of Nostradamus on its '*.ru 'web pages, but other countries have higher total ratios of term:page mentions than Russia. When ascribing a superranking encompassing the average of the approximate pages which mention Nostradamus and the average ratio of Nostradamus pages to total pages, Russia is the highest result. (Table 3)

Without removal of outliers, there is a significant Pearson's *r* of strong strength for the relationships between the ranks of raw HCEs for the average number of pages on a TLD and the 'average term per page' is r = 0.959065 (n = 38). This would suggest that the more pages on a domain, the more Nostradamus mentions it will have. The removal of outliers (top 5 superranks (6 total due to tie) and the lowest superrank) marginally improves the relationship to r = 0.997066 (n = 31). (Table 4)

As Figure 1 shows, Russia is considerably above the trend-line for the number of mentions of Nostradamus on pages, as are China, Japan, and Romania. Notably, of these, only Romania used a Latin alphabet and so it cannot be ruled out that there is some effect from non-ASCII characters on the accuracy of Google's HCEs. However, prior reporting has shown considerable interest in Nostradamus in Japan (Time 1999). Google Trends has shown prior high interest in Nostradamus in Ukraine and Romania (Hotchkiss 2017a, Hotchkiss 2017b).

6. Conclusion

Despite the need to expand the scope of the analysis to be truly generalizable, these findings support a view that Russia is an outlier among TLDs and news sites in terms of Nostradamus content, and that the content on the outlier Russian news sites is qualitatively different from typical Western coverage. This strongly supports prior research. Russia's state news agency RT promotes Nostradamus as a legitimate prophet-type figure (as does Sputnik News though it was not tested in this analysis. Sputnik hosts a significant number of Latin-alphabet pages in the Romanian language on Sputnik.md. The author's forthcoming research has highlighted RT content and Sputnik content which is related to Turkey as well. These findings are consistent with prior research and may support a view that Russia has greater cultural interest in Nostradamus than in the West, as well as a view that Russia may utilize Nostradamus prophecies in information operations and propaganda.

This research opens up some interesting questions as to why state news agencies like DW, RT and Sputnik may create Spanish-language Nostradamus content. Google Trends does show that South and Central America have some of the greatest interest in Nostradamus worldwide. This seems to be correlated to Catholic beliefs as well. It is also perplexing that the LA Times, which has run some of the most effective Nostradamus scepticism stories prior to 2001 (and especially in regards to earthquakes) would publish stories which advocate for Nostradamus' earthquake prediction skills (in Spanish).

There are many limitations to HCEs beyond just the fact that no search engine indexes the entire internet. In addition, there are many other variables which may affect the accuracy of HCEs which are returned for a given query, including the server from which the results are returned, the site from which the results are requested from (such as SEO effects from a specific domain), page to page variations in HCEs, biases in how web links may be indexed geographically, language or character set biases, server capacity planning effects, proprietary changes in search engine indexing algorithms over time, and potential effects of personalized search results, to name a few. In short, there are many potential deficiencies to the reliable utilization of HCEs for decision making. With that said, they represent a tantalizingly large and immediately available global sample population from which to conduct social science 'experiments'. Hopefully someday, accurate HCEs will be a reality, as they have much potential value in social science research.

It seems that the only way to even hope to extract good data from HCEs currently is to collect them over time to examine them for consistency and then examine the most relevant pages for qualitative context as was done here. Regardless, these results do not exist in a vacuum and without other background for comparison, are not useful.

References

Bjorneborn, L. & Ingwersen, P. (2004) *Toward a Basic Framework for Webometrics,* Journal of the American Society for Information Science and Technology. 55(14), pp. 1216-1227

Boghardt, T. (2006) Active Measures: The Russian Art of Disinformation*, AIRSHO magazine*, Oct. 2006, pp. 20-26.

Christensen, M. (1998) The Russian Idea of Apocalypse: Nikolai Berdyaev's Theory of Russian Cultural Apocalyptic, *Proceedings of the Third Annual International Conference of the Center for Millennial Studies*, Boston, MA, U.S.A.

Hotchkiss, M.B. (2017a) Russian Active Measures and September 11, 2001: Nostradamus Themed Disinformation?, *International Journal of Cyber Warfare and Terrorism (IJCWT),* 7(1), pp.25-41.

Hotchkiss, M.B. (2017b June 29-30) Nostradamus Prophecy as a Russian Information Warfare Concept. In M. Scanlon and N. Le-Khac. (editors). *Proceedings of the 16th European Conference on Cyber Warfare and Security. University College Dublin Ireland 29-30 June 2017*. pp. 172-175. Reading, UK: Academic Conferences and Publishing International Ltd.

U.S. Joint Publications Research Service (JPRS) (1978) *Translations on USSR Science and Technology: Biomedical and Behavioral Sciences,* Central Intelligence Agency (CIA), accessed at: https://www.cia.gov/library/readingroom/docs/CIA-RDP96-00787R000500430001-1.pdf

Rosenberger, L. & Berger, J.M. (02 August 2017) *Hamilton 68: A New Tool to Track Russian Disinformation on Twitter*. From Alliance for Securing Democracy (German Marshall Fund for the US) (blog). Available at:http://securingdemocracy.gmfus.org/blog/2017/08/02/hamilton-68-new-tool-track-russian-disinformation-twitter

Rousseau, R., (1999) Daily time series of common single word searches in AltaVista and NorthernLight. *Cybermetrics.* 2/3 (1998-9), Issue 1. Paper 2.

Uyar, A., (2009) Investigation of the Accuracy of Search Engine Hit Counts. *Journal of Information Science*. 35, 4, pp. 469-480.

Thelwall, M. (2008) Extracting Accurate and Complete Results from Search Engines: Case Study: Windows Live. *Journal for the American Society of Information Science and Technology,* 59(1), pp. 38-50.

Thelwall, M. 2009. *Introduction to Webometrics: Quantitative Web Research for the Social Sciences.* San Rafael, CA: Morgan and Claypool.

Thelwall, M., Vaughan, L., and Bjorneborn, L., (2006) *Webometrics.* Annual Review of Information Science and Technology, Pp. 81 – 135.

Time Magazine. (5 July 1999) Cover story: *Nostradamus Predicted that the World Would end this Summer: What are so many Japanese Taking him seriously?*, viewed at: http://content.time.com/time/covers/asia/0,16641,19990705,00.html

Wilson, I. (2007) *Nostradamus: The Man Behind the Prophecies.* Macmillan.