Googles Such-Einstellung gegenüber konservativen News-Seiten wurde quantifiziert

geschrieben von Andreas Demmig | 7. Oktober 2017

Abstrakt

Der Prozentsatz des Domainverkehrs, der von Google Search ausgewiesen wird, abgesehen von Recherchen nach Handelsmarken (PGSTN), liegt in oder um den Bereich 25% -30% für eine breite Klasse von Web-Domains. Diese Hypothese wird durch die Berechnung der Korrelation zwischen der Popularität von Webseiten für Nachrichten / Meinungen und ihrer PGSTN getestet, und befindet sich in der Nähe von Null. So kann PGSTN streng genutzt werden, um die vorsätzlich Suchalgorithmen der Google-Suche zu erkennen und sogar zu quantifizieren. Der Intentional Bias [die absichtliche bzw. vorsätzliche Ausrichtung – *ist dem Sinn nach doppelt gemoppelt, denn eine Ausrichtung / Einstellung ist immer absichtlich; daher gilt das auch für „Bias“ allein; Danke an unseren Leser Herrn Fred F. Müller – update vom 07.10.17*] ist der Bias, der durch interne Google-Vorgaben eingeführt wurde und unabhängig von externen Faktoren ist, wie die Dominanz [~ Häufigkeit] von bestimmten Standpunkten im Web. Hier wird die PGSTN-Methode für die absichtliche Bias-Erkennung bei der Klimadebatte und für den allgemeinen politischen Diskurs angewendet. Die Google-Suche ist äußerst voreingenommen für Klimaschutz und gegen den Realismus des Klimas. Die PGSTN-Bereiche für Klima-Realismus und Klima-Alarmismus überlappen sich nicht einmal! Einige der wichtigsten klima–realistischen Domains, darunter die nur wenig umstrittene *judithcurry .com* , haben eine so niedrige PGSTN, dass sie als auf der „schwarzen Liste“ von Google gelistet angesehen können. Es wurde herausgefunden, dass die Google-Suche links / liberale Domains bevorzugt und gegen konservative Domains eingestellt ist – mit einer Wahrscheinlichkeit von 95%. Weiterhin haben bestimmte Extrem-Links-Domains so ein hohes PGSTN, dass der Verdacht besteht, dass sie für eine prominente Platzierung handverlesen wurden. Bestimmte respektierte konservative Domänen sind auf der schwarzen Liste.

> [Hinweis zur Übersetzung:
>
> Für die Abkürzung: PGSTN habe ich nur weitschweifige Erklärungen gefunden. Es scheint mir, dass das die Abkürzung für bestimmte Verfahren / Begriffe in der Statistik ist. Im Sinn der Studie kann man es mit „Häufigkeit der Ergebnisanzeige“ gleichsetzen, oder wie im ersten Satz: „Prozentsatz des Domainverkehrs“ – Vielleicht weiß einer unserer Leser näheres.]

Einführung

Links-liberale politische Bias in der Google-Suche wurden seit Jahren bemerkt. Siehe Robert Epstein et al., *Eine Methode zur Erkennung von Bias in Such-Rankings, mit Nachweis der systematischen Bias im Zusammenhang mit der Präsidentschaftswahl 2016* ; Todd Dunning's *2015-09-15 Kommentar* ; Leo Goldstein, *Warum sind Suchmaschinen so feindlich gegen den Realismus?*
Diese Beobachtungen haben die Frage nicht vollständig gelöst, ob die Bias vorsätzlich so eingestellt sind oder den voreingenommenen Webinhalt widerspiegelte. In letzter Zeit zeigen die offiziellen *Suchbewertungsrichtlinien von* Google eine vorsätzliche Ablehnung gegen den Klimarealismus. Immerhin bestätigte ein ehemaliger Google-Mitarbeiter eine absichtliche Ablehnung von „allem Nicht-PC" [nicht-political correctness] durch das Google Search-Team.
Dieses Studie verwendet veröffentlichte SEO-Daten [Suchmaschinen Optimierungs-Daten] aus mehreren Quellen, darunter BrightEdge Research, *2017: Organische Suche ist immer noch der größte Kanal* und aktualisiert *seinen Bericht 2014* . Von nun an, bedeutet der Begriff " *Bias"* vorsätzliche Vorspannung. Diese Studie formuliert, begründet und wendet eine quantitative Methode der Bias-Erkennung in Google Search an.

Methoden

Es ist bekannt, dass die Google-Suche 25% -30% der Suchkriterien der [bisherigen] Nutzer für eine durchschnittliche Website bietet. Wie Googles Führungskräfte und die Google Werbung viele Male wiederholt feststellen, [soll] der Google Search Service die relevantesten und nützlichen Ergebnisse für die Abfragen des Benutzers liefern. Der Google-Vorsitzende Eric Schmidt scherzte sogar, dass es für jede Abfrage nur ein Ergebnis geben sollte – das Ergebnis, das der Benutzer sucht. Google-Server durchstreifen das gesamte Web, extrahieren Text, Links und andere Daten aus Billionen von Seiten. Google kämpft ständig und erfolgreich gegen Versuche, Websites durch abgekartete Verknüpfung und andere Suchmaschinen-Optimierung Techniken künstlich zu fördern.
In seinen Unternehmen verwendet Google auch eine enorme Menge an Off-Web-Informationen, die er durch Chrome-Browser, andere Google-Anwendungen und -Dienste, Analysen der„Leuchtfeuer", Domains-Registrier-Status und so weiter sammelt. Diese Informationen beinhalten die Popularität und Besitz der Domains. Google verarbeitet auch die spontane Rückmeldung von den Benutzern in Form von Häufigkeit der Klicks auf die Ergebnisse, Weiterklicks (~ zappen], die Häufigkeit wiederholter Recherchen mit geänderten Begriffen usw.
Google ist dabei sehr gut in seinem Job. Websites und Domains, die bei den Besuchern in der Regel weniger beliebt sind, werden als weniger wahrscheinlich eingeschätzt, um von Google angezeigt zu werden und umgekehrt. Der Effekt ist, dass der Prozentsatz der Daten-Recherche, den Domains von Google Search erhalten, über alle Web-Domains ähnlich ist.

Diese Tatsache wird dargestellt, durch die nahezu „null Korrelation“ zwischen der Popularität der Domänen und dem Prozentsatz des Netzverkehrs, den sie von Google in jedem der Sätze der linken / liberalen Medien und konservativen Medien erhalten, trotz der Domänen-Popularität (nach Alexa.com, niedrigere Werte bedeuten höhere Beliebtheit [~Rangfolge]), von 24 bis 1.469 für die links / liberalen Medien-Set, und von 56 bis 12.795 für die konservativen Medien. Der „Verkehr“ von Google-Anzeigen liegt bei etwa 5% des gesamten Google-Traffic, also ist es kein Faktor.
“Net Traffic“ [~Netzverkehr], wie in dieser Forschung verwendet, schließt den Netzverkehr aus, bei dem die Nutzer gezielt mit Namen suchen, wie *„foxnews“* oder *„fox news“* oder *„foxnews .com“*. Net Traffic reflektiert Googles Absicht besser, weil die Google-Suche nicht viel Auswahl hat, wenn der Benutzer nach einer Website mit ihrem Namen oder Marke sucht. www.Alexa.com bietet Informationen, die PGSTN-Berechnung für Hunderttausende von Web-Domains ermöglicht.
Angesichts der Robustheit der PGSTN, schließe ich, das die **statistisch signifikanten Unterschiede in PGSTN zwischen den a priori definierten Sätzen von vergleichbaren Domains auf absichtliche Bias von Google zurückzuführen ist**, es sei denn, es gibt eine andere gute Erklärung.

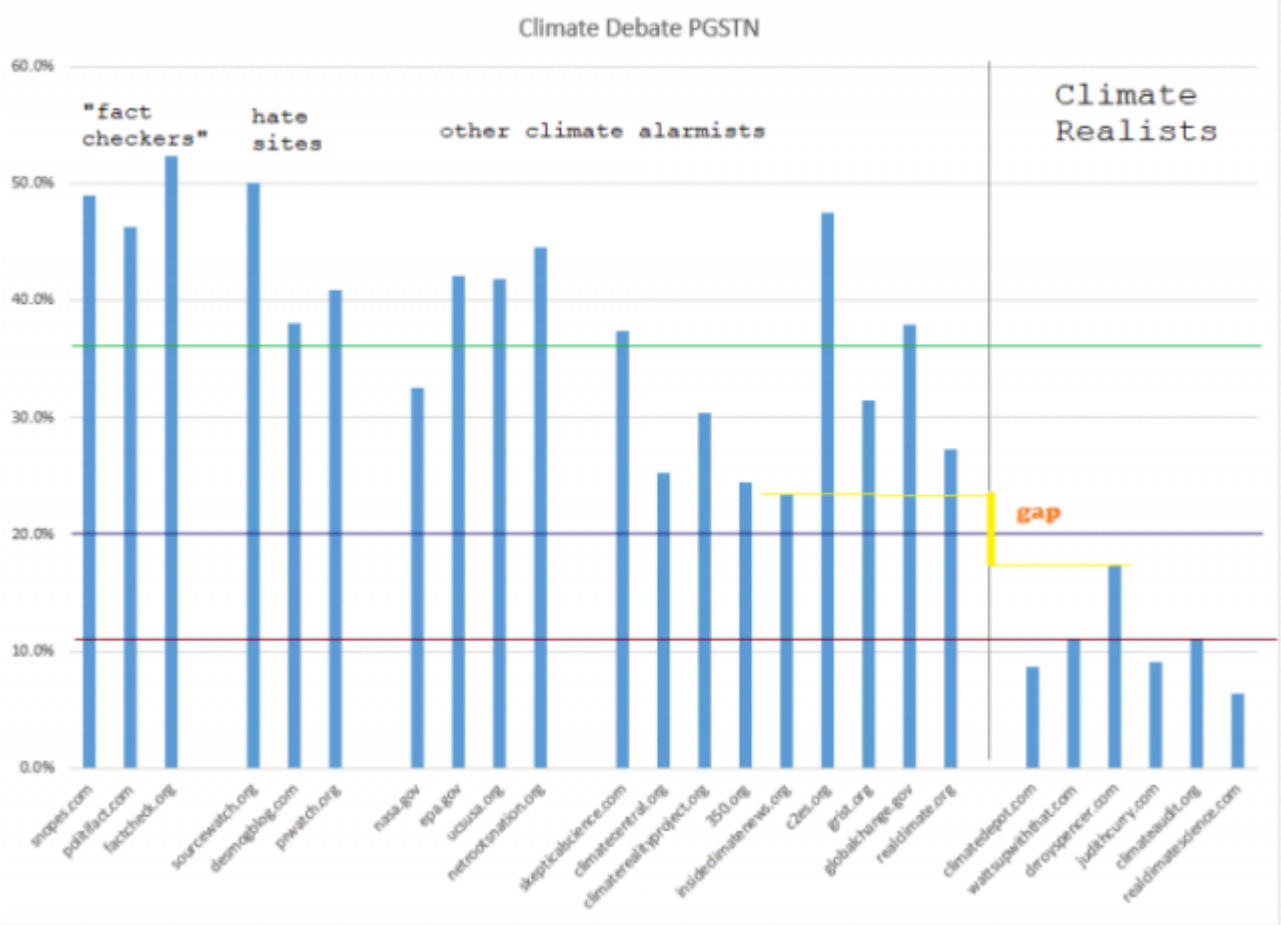
Methodische Details

Alle Daten in dieser Recherche basieren auf Schnappschüsse von Alexa (freie Version) vom 4. September 2017. Für jede Domain wurde ***Google Search Total*** aus der ***„Upstream Site“*** Tabelle entnommen – *Welche Seiten haben die Leute unmittelbar vor dieser Seite besucht*.
Die Anzahl der Recherchen nach Namen wurde von der Tabelle genommen ***„Top Keywords von Suchmaschinen“ – „****Welche Suchschlüsselwörter senden den Traffic auf diese Seite?* „.Es sollte beachtet werden, dass nur fünf Top-Werte, die in den freien Alexa-Schnappschüssen erscheinen, verwendet wurden. Alle in der Tabelle gezeigten Google-Suchdomänen wurden aufgenommen (*google.com, google.ca, google.de, google.co.in* etc.) Wenn die Gesamtmenge des Datenverkehrs nach Namen weniger als 5% betrug, wurde der Wert 1% eingegeben. PGSTN wurde durch Abzug von Namen-Suchverkehr aus dem gesamten Google-Suchverkehr berechnet.
Es wird nicht erwartet, das PGSTN ausreichende Sicherheit für einzelne Domains bietet, weil mehrere Faktoren das beeinflussen, einschließlich möglicher Fehler in den Alexa Daten. Trotzdem wurde die Google-Einstellung gegenüber einer Domain vorläufig in der beigefügten Kalkulationstabelle PGSTN-Domains.xlsx [Link zu einer Excel Mappe] wie folgt kodiert:
Whitelist / Grüne Zahlen:> 36%
Normal: 20% -36%
Graue Zahlen: 12% -20%
Blacklist: <= 12%
Es wurde erwartet, dass die meisten Domains (basierend auf der zitierten SEO-Forschung) PGSTN im Bereich von 20% -36% haben. Diese Erwartung wurde erfüllt. PGSTN <= 12% gibt vorläufig an, dass die Domain von

Google aufgeführt ist. Alles zwischen der schwarzen Liste und dem normalen Bereich gilt als grauer Bereich. Schließlich zeigt PGSTN> 36% vorläufig eine ungewöhnliche Bevorzugung durch Google an.

Google Bias in der Klimadebatte

Die Domains wurden meist nach Alexa-Klassifikation ausgewählt. Der Nachweis extremer Bias Google-Suche machte gegen Domains des Klima-Realismus keine statistischen Methoden erforderlich.



Häufigkeit der Anzeige bestimmter Webseiten als Ergebnis von Suchanfragen
Bezeichnung der Gruppen: Fakten Prüfer; Hassseiten, andere Klima Alarmisten, Klima Realisten
Es gibt eine große Lücke zwischen PGSTN von Realismus-Domains (6,3% – 17,4%) und PGSTN von Klima-Alarmismus-Domains (23,5% -52,4%). Die Lücke beträgt 6,1%. Mit Ausnahme von *drroyspencer.com* werden **alle Klima-Realismus-Domains von Google aufgeführt (PGSTN ist 6,3% – 11,0%)**.
Auf der anderen Seite haben selbst ernannte „fact *checker*„, darunter *snopes.com* und *politifact.com*, PGSTN von etwa 50%. Dieses schürt den Verdacht, dass sie von Google für die Priorisierung handverlesen worden waren. Weitere zwei Webseiten mit verdächtig hohem PGSTN sind *Quellwatch.org* (PGSTN = 50.1%) und *prwatch.org* (PGSTN = 40.9%). Diese beiden Seiten tauschen ihre Links gegenseitig aus (sie beziehen sich jeweils aufeinander als Quelle), haben überlappende Inhalte und es ist

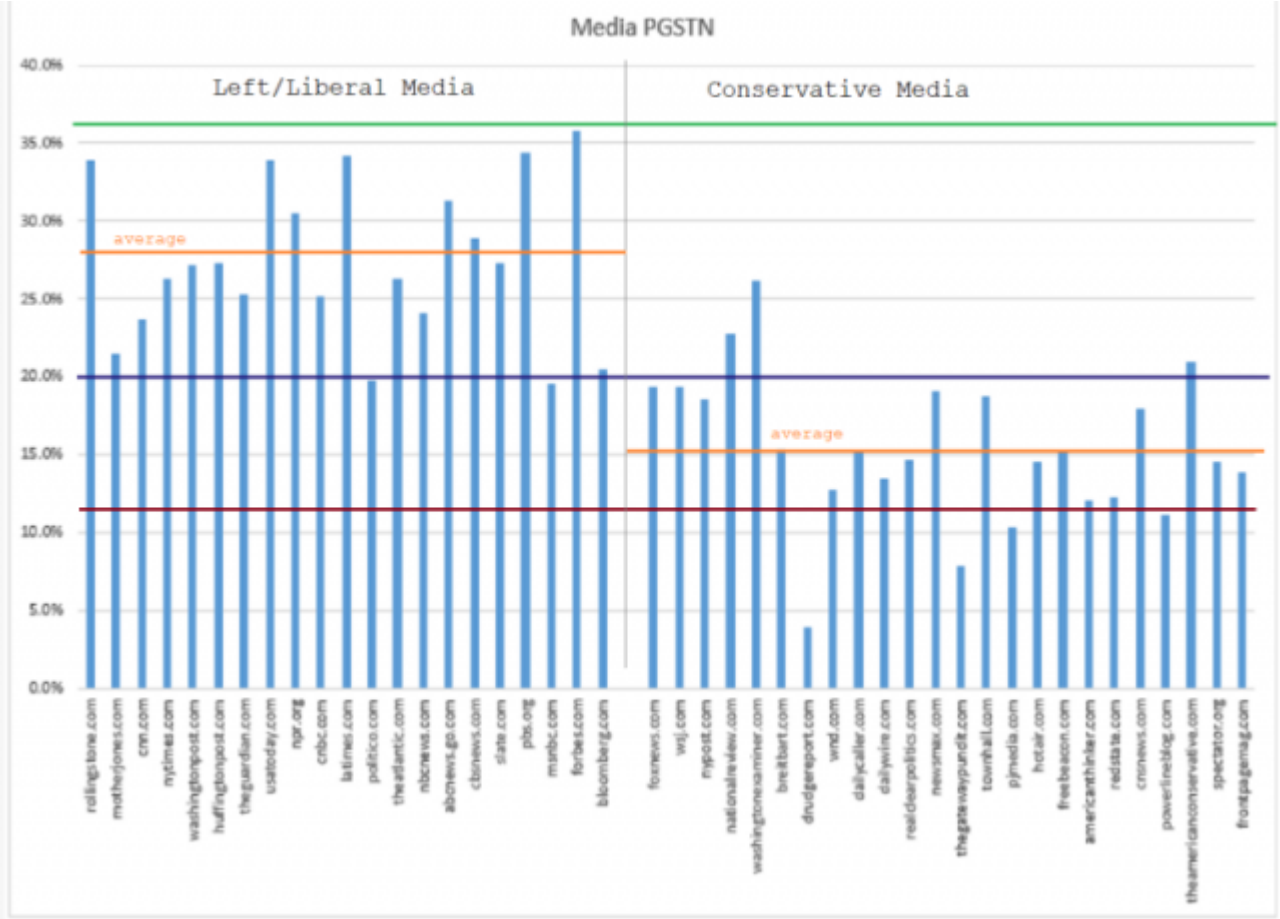
bekannt, dass sie der gleichen Organisation, dem *Center for Media and Democracy [Zentrum für Medien und Demokratie*] angehören. Dieses sind bekannte Anzeichen von Spam – doch Google hat es nicht nur geschafft, sie als Spam nicht erscheinen zu lassen, sondern sie sogar manuell zu priorisieren.
Dieser Abschnitt schießt *netrootsnation.org* mit ein, eine radikale linke Seite mit [Aufzeichnungen ~podcasts] verschiedener Tagungen, ist nicht speziell auf *Klimaalarmismus* ausgerichtet. Sein PGSTN = 44,5%. Diese Domain hätte handverlesen sein können oder ihre Besitzer wurden von Google-Insidern beim Spielen mit der Rangliste beraten. Google hat die Konferenzen finanziert, und die Google-Vertreter haben es besucht und Vorträge zu relevanten Themen gemacht, wie dieses – Ein Zitat:

> *„Wir werden einige Möglichkeiten teilen, um die Macht von Online-Videos zu nutzen und wie man Google und YouTube-Tools mit anderen Interessenvertretungen integrieren kann.“*

Alle anderen alarmistischen Domains haben PGSTN im Whitelist oder Normalbereich.

Google Bias im allgemeinen politischen Diskurs

Zur Quantifizierung von Google allgemeiner politischer Einstellungen, wählte ich Webseiten von Top-US-Nachrichten und Meinungsseiten nach ihrem Ranking in Alexa. Dann fügte ich einige konservativen Websites ein, mit niedrigerem Ranking basierend auf meinem persönlichen Wissen und / oder Alexa Vorschläge. Es gab ein Element der Subjektivität bei der Auswahl und Klassifizierung und ich habe einige Domains weggelassen, die ich nicht klassifizieren konnte. Trotzdem wurden die populärsten Domains sowohl in links / liberal (einschließlich Links, Mainstream Liberal und Mainstream Center) als auch in konservativen (einschließlich konservativen und Mainstream konservativen) Kategorien ausgewählt und eingestuft. Die Verwendung von gewichteten Statistiken minimiert das Element der Subjektivität in die Ergebnisse.



Die Ergebnisse zeigen, dass die **Google-Suche stark gegen konservative Domänen voreingenommen ist. Einige respektable konservative Domänen scheinen auf der schwarzen Liste zu sein;**
thegatewaypundit.com
pjmedia.com
americanthinker.com
redstate.com
powerlineblog.com
drudgereport.com
Es könnte eine alternative oder zusätzliche Erklärung für niedrige PGSTN des Drudge Report sein – diese Website besteht hauptsächlich aus Links zu Artikeln auf anderen Seiten, eine Praxis, die Google unterdrückt.
Im Durchschnitt erreichen die konservativen Domänen eine fast zwei Mal niedrigere PGSTN als die links / liberalen: konservative 15,5% (Standardabweichung 5,1%) gegenüber links / liberal 27,4% (Standardabweichung 4,9%). Die Hypothese der Google-Suche links / liberale Bias wird mit einer Wahrscheinlichkeit von 95% bestätigt.

Diskussion

Obwohl die PGSTN von einzelnen Domänen für Schlussfolgerungen nicht ausreicht, kann ich nicht vermeiden festzustellen, dass extremistische Webseiten wie *dailystormer.com* (PGSTN = 13,6%; existieren zum Zeitpunkt der Untersuchung nicht mehr) und *dailykos.com* (PGSTN = 20,2%) von Google gegenüber vielen konservativen und Klima-realistischen Domains bevorzugt

werden.

Schlussfolgerungen

Google-Suche ist zugunsten von links / liberalen Webseiten gegen konservative Webseiten voreingenommen und ist zugunsten des Klimaalarmismus gegen den Klimarealismus äußerst voreingenommen.

Offenlegung

Ich habe Short-Positionen in Google-Aktien [d.h., er spekuliert auf fallende Kurse]

Referenzen

Die Referenzen sind im Text des Artikels.

Zusatzmaterialien

Alexa Schnappschüsse sind abrufbar unter https://defyccc.com/data/PGSTN-Snapshots.7z (komprimiert mit 7-Zip).

Erschienen auf WUWT am 08.09.2017
Übersetzt durch Andreas Demmig
https://wattsupwiththat.com/2017/09/08/a-method-of-google-search-bias-quantification-and-its-application-in-climate-debate-and-general-political-discourse/