Annika Richterich

Google Trends: Using and Promoting Search Volume Indicators for Research

Abstract: *This paper discusses methodological research developments related to the web service Google Trends. It reflects on the implications of data evaluation based on search engine queries. Recent methodological developments in quantitative research design can be traced back to the establishment of search engines as main gateways to online content. While Google Inc. uses its own received web search queries in order to maintain more specific services, such as the epistemological surveillance platform Google Flu Trends, it also presents excerpts from its databases publicly in Google Trends. The service indicates, for example, how frequently a search-term has been entered in Google, and where this query can be geographically located. Information on actual search volumes is not provided, however. Recent studies have drawn on Google Trends in order to analyse relations between these search volume indications and developments such as stock market moves. What is presented to the public and used in most of these studies, however, are merely surrogates and indicators of the original web search logs and search engine queries, rather than the data itself. Such developments should be seen critically, since the original data are exclusively available to respective media companies and selected scientists. Google Trends is supposed to communicate transparency and openness. As a symbolic gesture, it implies that Google 'hands back' parts of the user-generated search engine data to the public. Applications such as Google (Flu) Trends are staged as philanthropic investment, but are only one out of the many data mining possibilities that are based on the users automatically paying their search engine queries with the data they leave behind.*

Einleitung: „Like Manna from Heaven"?

Google Trends erweist sich für den marktführenden Suchmaschinenanbieter *Google Inc.* als Balanceakt. Einerseits inszeniert sich das Unternehmen mit diesem Service auf gewohnte Weise transparent und offen. Unter den Kategorien ‚Hot Searches/Angesagte Suchanfragen' und ‚Top Charts' können Nutzer an Suchmaschinen-eingaben orientierte Trends tagesaktuell und thematisch gebündelt verfolgen. Unter ‚Explore/Erkunden' lässt sich die Entwicklung von

Suchbegriffshäufigkeiten geographisch sowie im zeitlichen Verlauf nachvollziehen. Damit wird suggestiv vermittelt: Die von Nutzern weltweit generierten Daten aus *Google*-Suchmaschineneingaben bleiben nicht nur kommerziellen Kunden und dem Unternehmen selbst vorbehalten, sondern werden zudem für private Nutzer zugänglich gemacht. Insofern lässt sich *Google Trends* auch als vermeintlich freigiebige Öffnung auffassen, mit der *Google Inc.* die von den Nutzern generierten Daten symbolisch an diese ‚zurückgibt'.

Andererseits erscheint der für Deutschland noch recht neue Service (verfügbar seit 2013) nicht zuletzt als ein riskanter Schritt, da *Google* mit dem *Trends*-Service auch den Umfang und die Auswertungsmöglichkeiten dahinter liegender Datenbanken andeutet. Aus Sicht des Unternehmens mag hier entscheidend gewesen sein, dass die Popularisierung kommerzieller *Google*-Services ohnehin die Frage nach der Speicherung und Verwertung von Nutzungsdaten aufwirft.

Oberflächlich betrachtet macht *Google Inc.* mehr Zugeständnisse denn je hinsichtlich der Transparenz von Datenspeicherung und -auswertung. Es sind jedoch nur begrenzte, beschwichtigende Einblicke, die den Nutzern gewährt werden. Die strategisch kommunizierte Öffnung eines Unternehmens, dessen Geschäftsmodell auf der kommerziellen Verwertung von Nutzerdaten besteht, ist nicht zuletzt eine symbolische Geste: Man täuscht in einer pragmatischen Nutzbarmachung darüber hinweg, dass nicht die Daten, sondern allenfalls ihre Stellvertreter und Indikatoren offen gelegt werden. *Google Trends* ist in diesem Sinne nicht nur ein Service, der *Googles* Suchmaschinendaten verwertet und zugänglich macht, sondern diese zugleich strategisch als Teil der unternehmerischen Öffentlichkeitsarbeit einsetzt.

In *Google Trends* ist es gerade diese Vermengung von augenscheinlicher Transparenz und einer limitierten, gewissermaßen zensierten Veröffentlichung von Daten-Stellvertretern, die darauf aufmerksam macht, wie viel unser Suchverhalten über uns preisgibt – und wie viel von diesem Wissen ein Unternehmen wie *Google Inc.* nicht mit seinen Nutzern teilt. Zwar werden Suchbegriffe entsprechend ihrer Eingabehäufigkeiten in ‚Angesagte Suchanfragen' gerankt, und bei der Eingabe von Begriffen im Bereich ‚Erkunden' erhalten die Nutzer übersichtliche Liniendiagramme, jedoch beinhaltet der Service keine konkreten Angaben zu Suchvolumina. Unter ‚Angesagte Suchanfragen' findet sich nur der Hinweis ‚mindestens x Suchanfragen', wobei x je nach

Suchbegriff zwischen 10.000 und mehreren Millionen schwanken kann. Die unter ‚Erkunden' erzeugten Liniendiagramme zeigen die Eingabehäufigkeit im historischen Verlauf auf einer relationalen Skala von 1 bis 100. *Google Trends* erscheint aus dieser Sicht als offensive Kommunikationsstrategie, die einen transparenten, gemeinnützigen Umgang mit Nutzerdaten suggeriert, und in dem, was der Service nicht zeigt, die Frage aufwirft: Wem gehören suchmaschinengenerierte Daten?

In diesem Zusammenhang erscheint insbesondere relevant, wie sich Wissenschaftler_innen zu *Google Trends* sowie damit einhergehenden ethischen Problemstellungen und methodischen Unsicherheiten positionieren. Bereits 2007 war im Editorial der *Nature* die heute für viele weitere Bereiche treffende Beobachtung zu lesen:

> For a certain sort of social scientist, the traffic patterns of millions of e-mails look like manna from heaven. Such data sets allow them to map formal and informal networks and pecking orders, to see how interactions affect an organization's function, and to watch these elements evolve over time. […] But for such research to flourish, it must engender that which it seeks to describe. And so it is encouraging that computational social scientists are trying to anticipate threats to trust that are implicit in their work. Any data on human subjects inevitably raise privacy issues [...], and the real risks of abuse of such data are difficult to quantify. (Nature Editorial Board 2007: 637/638)

Damit wird – auch im Anschluss an die zuvor gestellte Frage – deutlich, dass suchmaschinengenerierte Big Data für die Wissenschaft sowohl ethische bzw. netzpolitische als auch methodische Problemstellungen mit sich bringen. Angesichts eines Angebots wie *Google Trends* sieht man sich der offenen Problematik gegenüber, wie seitens wissenschaftlicher Disziplinen mit den von Nutzern durch Suchanfragen generierten Daten und ihrer zensierten, ‚teilöffentlichen' Aufbereitung durch *Google Inc.* (sowie andere Unternehmen) umzugehen ist.

Man kommt kaum umhin, hinter *Google Trends* vielversprechende sowie zugleich problematische Daten zu sehen, deren Erhebung und Aufbereitung von außen betrachtet nur begrenzt nachvollzogen werden können. Diese sind überdies in einen unternehmerischen Kontext eingebettet und werden nur kontrolliert freigegeben. So ergab sich etwa während eines Workshops, zu dem eine Arbeitsversion dieses Artikels vorgestellt wurde, bereits der Diskussionspunkt, inwiefern allein die

Thematisierung von *Google Trends* einen Diskurs und ein Agenda-Setting vorantreibt, das von dem Unternehmen vorbestimmt ist. Daher ist an dieser Stelle festzuhalten, dass es in dem vorliegenden Artikel nicht darum geht, Forschung anhand von *Google Trends* durchzuführen. Vielmehr soll exemplarisch diskutiert werden, welche methodischen, ethischen und netzpolitischen Implikationen die Forschung mit *Google*-Suchmaschineneingaben mit sich bringt. Dies wird unter anderem in Hinblick auf die bereits existierende Forschung beleuchtet, die auf *Google Trends* zurückgreift.

Für die medienkritische Auseinandersetzung mit den Potenzialen und Problemen der Erhebung und Auswertung von Suchmaschinen-nutzungsdaten ist ein Aspekt ‚transaktionaler Nutzer- und Nutzungsdaten' entscheidend: Während Manovich „transactional data" im Kontext von Big-Data-Forschung zunächst lose als „the traces of people's online behavior" (2011/2012) definiert und damit Daten aus „web searches, sensor data or cell phone records" (ebd.: 1) umfasst, macht er auch auf einen entscheidenden Unterschied in deren Qualität und Zugänglichkeit aufmerksam. In vielen Fällen ist der Zugriff auf die eigentlichen Daten den Unternehmen vorbehalten, innerhalb derer sie entstehen; sie kommen nur im Business-to-Business-Bereich zum Einsatz oder sind einzelnen, für die jeweiligen Firmen tätigen Wissenschaftlern vorbehalten. Solche Daten können nicht über die jeweiligen Programmierschnittstellen abgerufen werden, wie es etwa im Fall von Twitter, Reddit oder Flickr zum Teil durchaus möglich ist (vgl. dazu Manovich 2011/2012). Gewissermaßen als Kompromiss produzieren Unternehmen wie *Google Inc.* öffentlich verfügbare Stellvertreter und Indikatoren der zugrunde liegenden Daten, die etwa Häufigkeiten und Relationen indizieren, jedoch keine tatsächlichen Suchvolumina offenlegen. Bislang lassen sich die *Google-Trends*-Angaben allenfalls als CSV-Datei (Comma Separated Value) herunterladen und somit tabellarisch anzeigen. Auch anhand dieser Tabellen kann man jedoch keine genauen Suchvolumina ablesen. Seit 2007 kursieren Gerüchte, dass es eventuell eine *Google Trends API* (Application Programming Interface/Programmierschnittstelle) geben solle (siehe Millis 2007). Ironischerweise zeigte DuVander im Februar 2012 anhand einer *Google-Insights*-Analyse die steigende Nachfrage nach einer solchen Programmierschnittstelle. Dazu schrieb er zudem:

> What developers want, they find a way to get. There's a python library to extract the Google Trends data. Those after a Google Insights API aren't so lucky. The best I can find is a forum thread with scientists begging for data. One even offered Google $1,000 per year. Of course, it's dangerous to hang your hat on something unsupported. For those willing to walk out on a limb, we list 19 unofficial APIs. (DuVander 2012)

In diesem Artikel soll daher auch diskutiert werden, welcher Entstehungslogik suchmaschinengenerierte Daten unterliegen, bzw. inwiefern sich dieser Kontext bereits jeder externen Datenanalyse entzieht, und was dies für die entsprechende Forschung bedeutet. Zunächst stelle ich dazu einen stark verkürzten historischen Abriss darüber vor, in welcher Form *Google* die durch Suchmaschineneingaben entstandenen Daten bisher der Öffentlichkeit zur Verfügung stellte bzw. stellt. Anschließend skizziere ich einige Beispiele, wie diese Daten wiederum von Wissenschaft und Wirtschaft aufgegriffen wurden.

Aufgrund des begrenzten Umfangs dieses Artikels können die aufgeworfenen Fragen nicht abschließend beantwortet werden. Das Ziel ist vielmehr, deutlich zu machen, dass eine vermeintliche Öffnung von Unternehmen, deren Geschäftsmodell auf der Produktion und Weiterverwertung von transaktionalen Daten beruht, zumeist wenig mehr als eine symbolische Geste ist. Sie täuscht in einer pragmatischen Nutzbarmachung – sei es für private Recherchen oder wissenschaftliche Analysen – darüber hinweg, dass man es nicht unbedingt mit Daten, sondern allenfalls stellvertretenden Indikatoren zu tun hat. In diesem Sinne erfasst die vorliegende Untersuchung zwei zentrale Ebenen von ‚Datenkritik': einerseits die netzpolitische Problematik, dass Wissenschaft und vor allem Öffentlichkeit heute nur eingeschränkt Zugriff auf die von ihnen erzeugten Primärdaten haben, und andererseits methodologische Reflektionen zur Erhebungsrealität von unzugänglichen Big Data.

Google Trends: Ein kurzer Rückblick

Seit Mai 2006 (Mayer 2006) kommt *Google Trends* den Wunschträumen einiger quantitativer Markt- und Sozialforscher stetig näher. *Googles* Suchmaschinenanfragen werden hier in übersichtlicher Diagrammform präsentiert, sortiert nach geographischer Lokalisierung der Eingabe (vgl.

Abb. 1). Zudem werden sie, dies gehört zu den neueren Eigenschaften des *Google-Trends*-Service, täglich bzw. im Abstand weniger Stunden aktualisiert (für ‚Hot Searches'). Dabei sind es nicht nur oder vielleicht *gerade* nicht die Marketingabteilungen, die man mit diesem Angebot erreichen will. Mit Services wie *Google AdWords* oder *Analytics* scheint dieser Business-to-Business-Bereich längst abgedeckt. Vielmehr wendet man sich an die Öffentlichkeit und, wie der Artikel „Predicting the Present with Google Trends" (2011) der *Google*-Ökonomen Hal Varian (Chief Economist) und Hyunyoung Choi (Decision Support Engineering Analyst) zeigt: an die Wissenschaft.

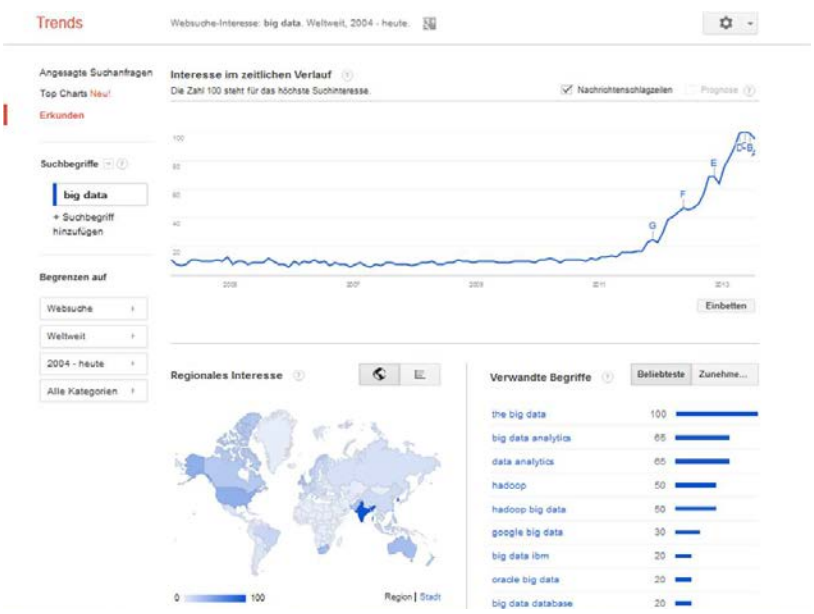


Abb. 1: Screenshot von *Google Trends* nach Eingabe von „Big Data"
(www.google.com/trends/explore, Screenshot vom 10.12.2013)

Google Insights for Search wurde Ende September 2012 in *Google Trends* integriert. Die Aufbereitung der Suchstatistiken veränderte sich jedoch nicht grundlegend. Abbildung 1 zeigt, wie *Google Trends* etwa die

Eingabehäufigkeit für den Suchbegriff „Big Data“ im zeitlichen Verlauf visualisiert. Die Maßzahl 100 steht weiterhin für die höchste Eingabehäufigkeit, ohne dass konkrete Suchvolumina angegeben werden. Zudem wird die Häufigkeit hier pauschal als Indikator für ‚Suchinteresse‘ veranschlagt. Diese diskursive Deutung ‚Interesse im zeitlichen Verlauf‘ ist bereits eine Verallgemeinerung von Nutzungsmotiven, die letztlich das datenkritische Problem fehlender Kontextualisierung aufwirft: Was genau wird mit diesen Häufigkeiten eigentlich qualitativ gemessen?

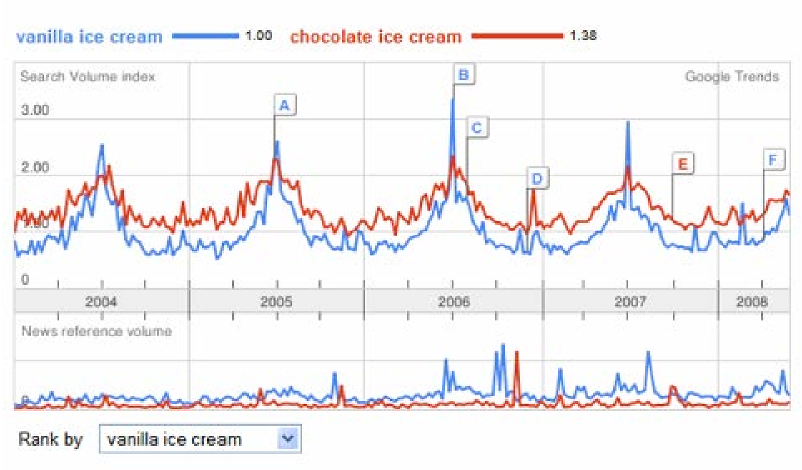


Abb. 2: *Google-Trends*-Visualisierung von Juni 2008 (Hwang 2008)

Nach der Demonstration einzelner Fälle, anhand derer Varian und Choi exemplarisch Forschungsdesigns vorführen und die Korrelationen von *Google-Trends*-Indikatoren und wirtschaftlichen Entwicklungen nahelegen, geben sie ihrer Hoffnung Ausdruck: „that these examples will encourage other researchers to experiment [with] this data source in their own research“ (2011: 9).

Wie bereits eingangs angedeutet, wird hieran offenkundig, dass die Veröffentlichung von *Google Trends* unter anderem der Referenzierung und Popularisierung dieser Datenquelle in entsprechenden akademischen Forschungsarbeiten dient. Hier sollte jedoch einschränkend erwähnt werden, dass es sich in diesem Fall offenkundig um anwendungsorientierte Forschung (Research & Development) handelt und zwischen unterschiedlichen Ausrichtungen von Forschung differenziert werden muss.

Bereits im Jahr 2001 veröffentlichte *Google Inc.* den sogenannten *Zeitgeist*-Service: Seitdem werden jeweils zum Jahresende „search patterns, trends and surprises“ (Schnitt 2001) vorgestellt. Der Service präsentiert ‚Hitlisten‘ der beliebtesten Suchanfragen des vergangenen Jahres, gruppiert nach thematischen Schwerpunkten. Dabei werden die thematischen Schwerpunkte in ‚Diagrammen‘ visualisiert, die keine numerischen Angaben enthalten. *Google Zeitgeist* sei ein „unique window into what is happening in the world on any given day, as well as a fascinating retrospective on the peaks and valleys of popular culture“ (ebd.).

2004 stellte das Unternehmen diesem jährlichen Service dann *Google Trends* zur Seite, das zunächst nur für die USA und in vergleichsweise rudimentärer Form verfügbar war. Abbildung 2 zeigt einen Vergleich zwischen Eiscremesorten, den *Google* im Juni 2008 in einer Pressemitteilung veröffentlichte. Anhand der y-Achse wird deutlich, dass auch hier keine Daten zu Suchvolumina zur Verfügung gestellt wurden. Im Gegensatz zum *Zeitgeist*-Format erhalten Nutzer im *Trends*-Service jedoch einen Zahlenwert, der zumindest einen relationalen Vergleich ermöglicht. Dazu ist in der Pressemitteilung auf dem offiziellen *Google*-Blog zu lesen:

> You’ll notice a number at the top of the graph as well as on the y-axis of the graph itself. These numbers don’t refer to exact search-volume figures. Instead, in the same way that a map might “scale” to a certain size, Google Trends scales the first term you’ve entered so that its average search volume is 1.00 in the chosen time period. So in the example above, 1.00 is the average search volume of vanilla ice cream from 2004 to present. We can then see a spike in mid-2006 which crosses the 3.00 line, indicating that search traffic is approximately 3 times the average for all years. (Hwang 2008)

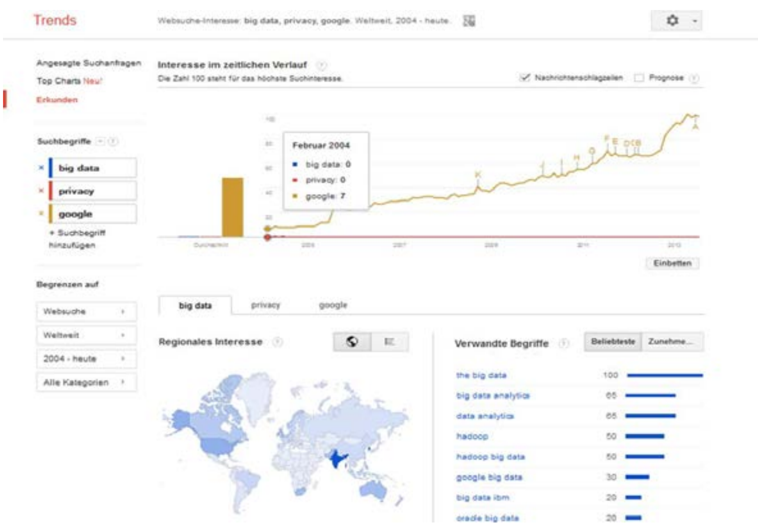
Auch die zuvor erwähnte Option, eine CVS-Datei der Anfrage herunterzuladen, war zu diesem Zeitpunkt (10. Juni 2008) neu und wurde in der Pressemitteilung bekannt gegeben.

Der im August 2008 eingeführte Service *Google Insights for Search* war im Vergleich dazu informatorisch reichhaltiger, wenngleich auch nicht quantitativ genauer. Hier fand man etwa geographische Visualisierungen, wie sie heute auch bei *Google Trends* verfügbar sind. Ähnlich der aktuellen Version von *Google Trends* (Stand: September 2013) bildet die y-Achse einen relationalen Vergleich ab, der den Höchstwert mit 100 ansetzt und davon ausgehend die anderen Werte bemisst.

Eine weitere, bereits angesprochene problematische Eigenschaft von *Google Trends* sind die relationalen Verschiebungen, welche bei der Eingabe mehrerer Suchworte auftreten. Gibt man etwa, so wie in Abbildung 3 gezeigt, „Big Data, Privacy“ ein, verändern sich die numerischen Angaben der einzelnen Suchbegriffe aufgrund der komparativen Eingabe. Über tatsächliche Größenordnungen lässt sich somit nur mutmaßen.

Noch deutlicher wird das anhand von Suchbegriffen, die eine vergleichsweise erhöhte Suchpopularität vermuten lassen, wie man z.B. anhand der Suchbegriffe „Big Data, Privacy, Google“ sieht (vgl. Abb. 4). Letztere Werte fallen im Vergleich nur noch marginal aus. Ein Grund dafür ist die Normalisierung der Daten, die *Google Inc.* im ‚Support‘ von Google Trends wie folgt kommentiert:

> Werden die Daten normalisiert? Ja. Alle Ergebnisse bei Google Trends sind normalisiert. Die Datensätze werden hierfür durch eine gemeinsame Variable geteilt, um die Auswirkung dieser Variable auf die Daten aufzuheben. Dadurch können die eigentlichen Merkmale der Datensätze miteinander verglichen werden. Bei der Anzeige der absoluten Rangfolge ohne eine Normalisierung der Ergebnisse würden die Daten aus Regionen mit dem höchsten Suchvolumen immer hoch eingestuft werden. Die folgenden Beispiele veranschaulichen einige entscheidende Punkte der Normalisierung:
>
> Kanada und Fidschi haben dieselben Prozentwerte für den Begriff ‚Hotel‘. Haben dadurch beide dasselbe Suchvolumen für diesen Begriff? Nein. Nur



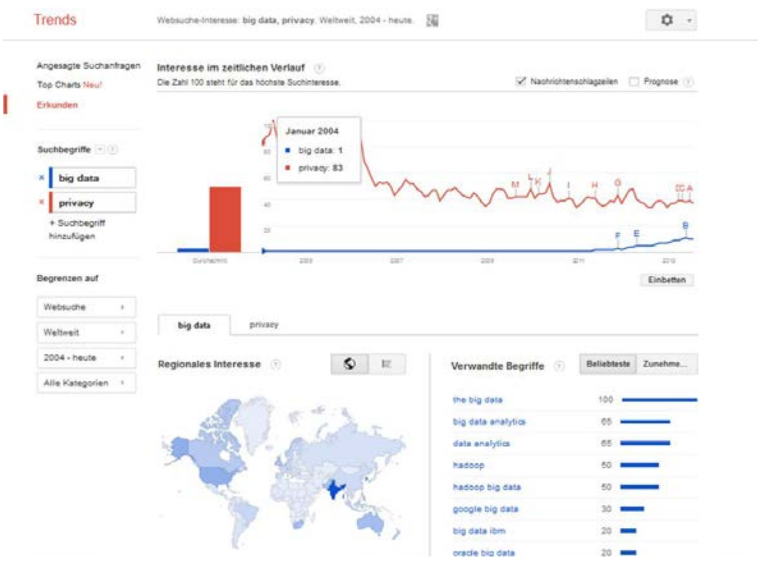


Abb. 3 und 4: *Google-Trends*-Visualisierung nach Eingabe von „big data, privacy“ und „big data, privacy, google“ (www.google.com/trends/explore, Screenshots vom 10.10.2013).

> weil zwei Regionen denselben Prozentwert für einen bestimmten Begriff haben, muss das absolute Suchvolumen nicht für beide gleich sein. Beide Regionen haben sehr unterschiedliche Suchvolumen. Die Daten können aber gleichwertig miteinander verglichen werden, da sie mithilfe des gesamten Traffics der jeweiligen Region normalisiert wurden. (Google Trends Support 2013)

Während *Google* die *Trends*-Daten zunächst nur jährlich, dann wöchentlich und immer häufiger aktualisierte, werden die zugrunde liegenden Daten mittlerweile nahezu in Echtzeit aktualisiert: Unter „Angesagte Suchanfragen“ findet man den Hinweis: „Vor etwa x Minuten aktualisiert” bzw. „Vor etwa einer Stunde/x Stunden aktualisiert”. Seit September 2007 aktualisiert *Google Inc.* die Daten des *Trends*-Services „Erkunden“ für die USA täglich (zwischenzeitlich wurden die Angaben sogar ebenfalls stündlich aktualisiert, vgl. Heisler 2008); seit Juni 2013 wird dies auch für Deutschland angeboten. Noch immer sind es jedoch nur zensierte, eingeschränkte Einblicke, die den Nutzern gewährt werden. Aus einer datenkritischen Perspektive ist folglich zu bedenken, wie die veröffentlichten Stellvertreter der Sucheingaben-Daten bereits von *Google* ‚zugerichtet‘ werden, und was dies nicht zuletzt für ihre Verwendung in wissenschaftlichen Kontexten bedeutet.

Zur Aus- und Verwertung von Google-Suchmaschineneingaben

Bei der wissenschaftlichen Auswertung von *Google*-Suchmaschinendaten lassen sich vor allem zwei Konstellationen beschreiben. Zum einen beschäftigt *Google Inc.* selbst Wissenschaftler_innen, die Zugriff auf die Suchmaschinendaten haben. Überdies kooperiert das Unternehmen mit ausgewählten Wissenschaftler_innen an Universitäten und anderen staatlichen Institutionen, denen es ebenfalls privilegierten Zugriff auf (u.a.) seine Web Search Logs gewährt. Zum anderen liegen mittlerweile verschiedene Studien vor, die ausschließlich die öffentlich zugänglichen Daten (in Form der CVS-Dateien) von *Google Trends* für unterschiedliche Auswertungen verwendeten. Für derartige Studien bleibt der Prozess der Datenerhebung und Normalisierung jedoch letztlich eine Blackbox. Im

Folgenden werde ich Beispiele für beide Formen der Forschung mit *Google*-Suchmaschinendaten näher vorstellen und diskutieren.

Als *Google Inc.* und die *US Centers for Disease Control and Prevention* (CDC) im Jahr 2008 erstmals *Google Flu Trends* publik machten, erhielt man einen Vorgeschmack der Verwertungsmöglichkeiten von *Google*-Suchanfragendaten. *Google Flu Trends* basiert letztlich auf (thematisch spezifischeren) Daten, wie sie in *Google Trends* in zensierter, normalisierter Form bereitgestellt werden, verbindet diese jedoch zusätzlich mit externen Datenquellen.[1] Der Service ist somit Resultat und Beispiel eines Studiendesigns, innerhalb dessen Wissenschaftler_innen einen privilegierten Zugang zu *Googles* Web Search Logs hatten: Zur Überwachung von Influenza-Intensitäten entwickelten Ginsberg u.a. eine Suchanfragen-Datenbank, die bisherige Eingaben mit aktuellen Anfragen abgleicht und deren Häufigkeit auswertet: „For the purpose of our database, a search query is a complete, exact sequence of terms issued by a Google search user […] Our database of queries contains 50 million of the most common search queries […].“ (Ginsberg u.a. 2009: 1014) Ursprünglich wurde diese semantische Grundlage sowie ihre Korrelation zu Influenza-Daten der CDC aus „hundreds of billions of individual searches from 5 years [2003-2008] of Google web search logs“ (ebd.: 1012) gewonnen.[2]

Google Flu Trends basiert auf einer Korrelation zwischen der Eingabe von Suchbegriffen, die sich mit Influenza in Verbindung bringen lassen, und der Feststellung von Influenza-Intensitäten in bestimmten Regionen. Als Basis des Projekts diente ursprünglich eine Studie, die für den Zeitraum von 2003 bis 2008 eine positive Korrelation zwischen Suchbegriffvolumina und medizinisch validierten Influenza-Statistiken feststellen konnte. Die Autoren kamen so zu dem Schluss, dass „the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-

---

[1] Damit ähnelt es strukturell *Google Correlate* (http://www.google.com/trends/correlate), einem Service, der es Nutzern erlaubt, Suchbegriffshäufigkeiten mit wöchentlichen/monatlichen Schwankungen oder der geographischen Lokalisierung von Suchmaschineneingaben (in US-Staaten) in Verbindung zu setzen.

[2] Bereits zuvor legten etwa Eysenbach (2002, 2006, 2009) und Polgreen u.a. (2008; *Yahoo Research*) Studien vor, die die Auswertungsmöglichkeiten von Web Search Logs nahelegten.

like symptoms“ (Ginsberg u.a. 2009: 1012).[3] Somit erlaube *Google Flu Trends*, so Ginsberg u.a., eine gezielte Auswertung von „Google search data to estimate current flu activity around the world in near real-time“ (Google Flu Trends 2011).

Einer solchen Analyse liegen Übereinstimmungen zwischen dem Ort der Eingabe und der geographischen Lokalisierung einer Infektionsverbreitung zugrunde; zudem ist eine Auswertung mit nur geringer zeitlicher Verzögerung möglich. In ihrer Untersuchung „Detecting influenza epidemics using search engine query data“ (2009) weisen die Autoren darauf hin, dass etwa Einrichtungen wie die US *Centers for Disease Control and Prevention* Grippeintensitäten lediglich mit einer ein- bis zweiwöchigen Verzögerung evaluieren können, während die von ihnen beschriebene, auf *Google*-Suchmaschineneingaben basierende Methode nur einem „reporting lag of one day“ (Ginsberg u.a. 2009: 1012) unterliege. Es handelt sich hier somit um Daten, die – infolge einer anfänglichen Validierung der Signifikanz spezifischer Stichwörter – dazu dienen sollen, Aussagen über Entwicklungen von Grippeintensitäten in Echtzeit zu machen und möglichst verlässliche Prognosen in Aussicht zu stellen.

Die Annahmen über die Verlässlichkeit von *Google Flu Trends* sind jedoch nicht gänzlich unproblematisch. Zwar hat der Service zunächst eindrucksvolle Übereinstimmungen zwischen Prognosen und tatsächlichen Epidemie-Intensitäten gezeigt. Dies erweckte den Anschein, dass es traditionellen Frühwarnnetzwerken, die direkt an ärztliche Berichterstattungen gekoppelt waren, überlegen sei. Jedoch wiesen Ginsberg u.a. in ihrem Artikel bereits auf eine Problematik hin,

---

[3] Die Autoren bezogen sich vor allem auf zwei Datenquellen: Sie nutzten sowohl online verfügbare Daten der *Centers for Disease Control and Prevention* für neun Surveillance-Regionen in den USA, als auch bundesstaatliche (und somit geographisch begrenzte) Daten für Utah. Unter http://www.cdc.gov/flu/weekly dokumentieren die CDC den durchschnittlichen Anteil ambulanter Patienten, die von Teilnehmern des *US Sentinel Provider Surveillance Networks* mit Influenza bzw. ‚influenza-like illness‘ (ILI) diagnostiziert wurden. Diese Daten werden wöchentlich innerhalb der jährlichen Influenza-Saison aktualisiert. Indem die Häufigkeiten verschiedener Suchanfrageneingaben zu diesen Daten traditioneller Influenza-Surveillance in Beziehung gesetzt wurden, konnte eine Korrelation zwischen der Suchanfragenentwicklung und den Influenza-Aktivitäten in den USA festgestellt und validiert werden.

die seither mehrfach zum Tragen kam: „[P]anic and concern among healthy individuals may cause a surge in the ILI-related [influenza-like illness, Ergänzung AR] query fraction and exaggerated estimates of the ongoing ILI percentage“ (ebd.).

Wie Butler in seinem Artikel „When Google got flu wrong“ (2013) herausstellte, waren jüngst jedoch erhebliche systematische Schwächen in den Algorithmen festzustellen, die den *Google-Flu-Trends*-Prognosen zugrunde lagen. Anfang 2013 überschätzten die Prognosen von *Google Flu Trends* die Grippeepidemie-Wahrscheinlichkeit in verschiedenen Regionen der USA bei weitem:

> [T]he latest US flu season seems to have confounded its algorithms. Its estimate for the Christmas national peak of flu is almost double the CDC's (see ‚Fever peaks'), and some of its state data show even larger discrepancies. It is not the first time that a flu season has tripped Google up. In 2009, Flu Trends had to tweak its algorithms after its models badly underestimated ILI in the United States at the start of the H1N1 (swine flu) pandemic — a glitch attributed to changes in people's search behaviour as a result of the exceptional nature of the pandemic. (Butler 2013; vergleiche zu dieser Problematik auch Cook u.a. 2011)

Es zeigt sich, dass *Google Flu Trends* vor allem in von bisherigen Standards abweichenden Fällen nicht konstant verlässlich ist. Stattdessen erfordert der Service eine kontinuierliche Evaluierung und Anpassung. Da das Angebot auf historischen Mustern aufbaut, wird jede Abweichung vom bisherigen Suchverhalten und daran orientierten Korrelationen zum Störfaktor. Brownstein, der am Crowdsourcing-basierten Influenza-Projekt Flu Near You beteiligt ist, kommentierte *Googles* Fehlkalkulation: „You need to be constantly adapting these models, they don't work in a vacuum […]. You need to recalibrate them every year.“ (Butler 2013). Bereits im Jahr 2006 verwies Eysenbach darauf, dass „Epidemics of Fear“ (Eysenbach 2006: 244) die Verlässlichkeit gesundheitsbezogener Suchmaschinendaten und die Stabilität angenommener Korrelationen beeinträchtigen können.

Vor allem aber basiert *Google Flu Trends* nicht auf ärztlichen Diagnosen oder virologischer Diagnostik von Influenza oder Todesfallstatistiken, wie sie traditionelle Methoden epidemiologischer Surveillance verwenden. Auch muss eine grippebezogene Suchanfrage nicht notwendigerweise von eigenen oder im Umfeld festgestellten Symptomen motiviert sein. Damit bleibt letztlich zu Teilen unklar, was die Nutzer bzw. ihre Sucheingabe (außer einer Erkrankung) motivieren könnte. Auch der mediale Diskurs, Zeitungen, Fernsehen, Onlinequellen, und die Berichterstattungen zu Epidemien, selbst in geographisch entfernten Regionen, können das Suchverhalten beeinflussen. So schreibt auch Butler, dass die Abweichungen zwischen der Prognose von *Google Flu Trends* und den tatsächlichen Grippeintensitäten unter anderem mit starken regionalen Differenzen und der verstärkten Berichterstattung zusammenhingen:

> [S]everal researchers suggest that the problems may be due to widespread media coverage of this year's severe US flu season, including the declaration of a public health emergency by New York State last month. The press reports may have triggered many flu-related searches by people who were not ill. (Butler 2013)

Fehlprognosen werden folglich dann wahrscheinlich, wenn externe, neue Faktoren zu Beweggründen für Sucheingaben werden, die von den bisher implizit in den Algorithmen des Prognosesystems eingeschriebenen Motivationen abweichen. Insofern muss ein Projekt wie *Google Flu Trends* die Deutung bestimmter Suchbegriffe kontinuierlich an historische Muster anpassen.

Störfaktoren können somit einerseits unvorhergesehene Ereignisse oder Diskurse sein, die das Suchverhalten der Nutzer verändern. Man sollte zudem auch eine gewisse ‚Selbstreferenzialität' des Systems in Betracht ziehen: *Google Flu Trends* kann vor allem dann funktionieren, wenn die Veröffentlichung dieser Daten nicht die Motive der Nutzer beeinflusst, auf deren ‚natürlichem' Verhalten die Datengenerierung aufbaut.[4] (Es sei denn, der Algorithmus ist in der Lage, dies

[4] *Google Trends* könnte demgegenüber von einer Verstärkung der als ‚Angesagte Suchanfragen' oder ‚Top Charts' gekennzeichneten Anfragen (aus unternehmerischer Sicht) profitieren, da die als Trends indizierten Ergebnisse von den Nutzern ggf. wiederholt werden. In *Google Trends* kann etwa in den beiden genannten Kategorien die Suchanfrage mit einem Klick auf den jeweiligen Begriff wiederholt werden. Hier ist unklar, ob und wie dies von *Google* wiederum in die Ergebnisse mit einbezogen wird. Vor allem

einzukalkulieren.) *Google Flu Trends* ist somit ein Fallbeispiel, anhand dessen sich veranschaulichen lässt, wie *Google Inc.* die analytische Aussagefähigkeit seiner Suchmaschinendaten kontinuierlich mit externen Daten (re-)validiert und – durch die thematische Wahl – öffentlichkeitswirksam einsetzt. Dem Service liegen konkrete Suchvolumina zugrunde, die jedoch nicht öffentlich gemacht werden. In ähnlicher Weise, wenngleich thematisch offener, zeigt auch *Google Trends* Sucheingabenhäufigkeiten im zeitlichen und geographischen Vergleich, ohne dabei jedoch tatsächliche Suchvolumina offen zu legen. Im Folgenden wird beispielhaft eine Studie betrachtet, die auf den limitierten Angaben von *Google Trends* basiert.

Neben Einsatzversuchen in der Epidemiologie besteht seit mehreren Jahren ein gesteigertes Interesse daran, Suchmaschinendaten auch für Wirtschafts-/Börsenprognosen nutzbar zu machen. Ein aktueller Artikel von Preis, Moat und Stanley (2013) demonstriert, dass *Google Trends* vor allem aufgrund der mittlerweile täglich erfolgenden Aktualisierungen aussagefähig sein können. Ihr Ansatz ist Beispiel für ein Forschungsdesign, das keinen privilegierten Zugang zu tatsächlichen Suchvolumina hat, sondern nur auf die öffentlich zugänglichen ‚Stellvertreter'/Normalisierungen zugreift. Anstelle einer Korrelation von Begriffen, die mit Gesundheitsfragen bzw. Influenza in Verbindung stehen, präsentierten die Autoren Zusammenhänge zwischen *Google*-Suchmaschineneingaben, wie z.B. dem Begriff ‚debt', und der Entwicklung des *Dow Jones Industrial Average*. Ihre Analyse basierte auf Daten, die von dem *Google-Trends*-Service bereitgestellt wurden. Entgegen bisheriger Studien, die sich zum prognostischen Potential von *Google Trends* skeptisch verhielten (vgl. dazu Preis/Reith/Stanley 2010), verwendeten die Autoren täglich aktualisierte Echtzeitdaten des Ende 2012 aktualisierten *Google-Trends*-Services. Wie David Leinweber in seinem Forbes-Artikel „Big Data gets Bigger" (2013) herausstellte, hängen aktuelle Feststellungen von positiven Korrelationen u.a. mit der geringeren Verzögerung zwischen Datenerhebung und Ereignissen zusammen.

Eine Analyse der quantitativen Veränderungen in *Google*-Suchanfragen, welche die Autoren (zunächst potentiell) als

---

für kommerzielle Anbieter erscheint *Google Trends* damit auch als Plattform, die es strategisch zu besetzen gilt.

finanzrelevant klassifizierten, eröffnet den Blick auf „patterns that may be interpreted as ‚early warning signs' of stock market moves" (Preis/Moat/Stanley 2013). Auf dieser Grundlage demonstrieren sie hypothetisch die Profitabilität einer Investmentstrategie, die *Google-Trends*-Daten als Entscheidungsgrundlage nutzte. Ihre Entscheidungen, Aktien zu kaufen oder zu verkaufen, beruhten auf der Ab- und Zunahme spezifischer Sucheingaben. So hätte etwa eine Investmentstrategie basierend auf dem Begriff ‚debt' einen Gewinn von 326% eingebracht. Ein solcher Erfolg validiert die anfängliche Annahme der Autoren, dass ein Anstieg im Wert des *Dow Jones Industrial Average* auf eine Verminderung der *Google*-Suchanfragen zu bestimmten finanzrelevanten Begriffen (wie ‚debt') folgte. Zugleich ging ein Anstieg bestimmter Suchbegriffshäufigkeiten mit einem Wertverlust des DJIA einher. Inwiefern diese Korrelation jedoch historisch spezifisch ist, bleibt offen. Angesichts der Börsenkrise 2011 kommt dem Begriff ‚debt' eine situativ hohe Bedeutung zu, die sich nicht unbedingt als stabil erweisen muss. Es erscheint naheliegend, dass nur eingeschränkt von einem prognostischen Potential ausgegangen werden kann, da gerade Begriffe wie ‚debt' eine historische Dimension nahelegen und die Häufigkeit als Suchbegriff von vielfältigen Diskursen und möglichen Beweggründen beeinflusst wird.[5]

Dies führt erneut zu der Problemstellung zurück, dass die Korrelationen zwischen bestimmten Suchbegriffen und prognostizierten Entwicklungen auf einer Deutung beruhen: Welchem Beweggrund folgte eine Mehrheit der Nutzer für die Eingabe eines bestimmten Begriffes bzw. inwiefern reflektieren solche Eingaben ein spezifisches Wirtschafsklima? Damit stellt sich erneut die Frage nach der Fluidität des

5 Neben ‚debt' experimentierte man auch mit anderen Schlüsselbegriffen (insgesamt 98), um deren Einfluss auf die beschriebenen Kauf-/Verkaufsstrategien auswerten zu können: „Cumulative returns of 98 investment strategies based on search volumes restricted to search requests of users located in the United States for different search terms, displayed for the entire time period of our study from 5 January 2004 until 22 February 2011 – the time period for which Google Trends provides data" (ebd.). Während ‚inflation' ‚housing' oder ‚bonds' eine signifikante Korrelation zu Börsenentwicklungen nahelegen, verfügten die Begriffe ‚freedom' ‚home' oder ‚labor' in dieser Hinsicht über kein aussagefähiges Potential. Die Studie der Autoren lässt sich insofern auch als Entschlüsselung eines semantischen Netzes von Börsenentwicklungen lesen. Mehrdeutige Begriffe erwiesen sich hier abermals als Herausforderung.

Nutzungskontexts und der Motive für die Eingabe bestimmter Suchbegriffe. Situative, kontextuell bedingte und damit unkontrollierte Anpassungen des Nutzungsverhaltens können zu Veränderungen in allgemeinen Suchmustern führen. Damit geht einher, dass Aussagen, die von dem Ergebnis eines bestimmten Algorithmus abgeleitet werden, nicht länger zutreffen. Die Algorithmen müssen also kontinuierlich an sich ändernde Deutungsmuster und Kontextualisierungen von Sucheingaben angepasst werden.

Dabei kann es sich letztlich nicht nur um unabsichtliche Änderungen des Nutzungsverhaltens handeln, sondern auch gezielte Anpassungen des Nutzungsverhaltens können eine solche Fluidität verursachen. Abweichungen im Nutzungsverhalten sind aus verschiedenen Gründen denkbar. Im Fall von *Google Flu Trends* waren es abweichende Motive in Hinblick auf Influenza-Intensitäten, die zu einer Störung bisheriger Korrelation führten. Angesichts jüngster Datenschutzskandale wird es zudem mehr denn je wahrscheinlich, dass Nutzer ihr Suchmaschinenverhalten bewusster steuern und limitieren. Vorwürfe, wie sie die Weitergabe von Nutzerdaten führender Internetkonzerne an das Prism-Programm der US-amerikanischen *National Security Agency* (*NSA*) betreffen, sowie der Skandal um das Tempora-Programm des britischen Government Communications Headquarters (GCHQ), werfen die Frage auf, inwieweit sich das Nutzerverhalten vermehrt einem Bewusstsein um das digitale Panoptikum anpasst. Während sich bislang die Grenze zwischen ‚Verschwörungstheorien' und Datenausspähung noch diskursiv aufrechterhalten ließ, bieten die neuesten Enthüllungen kaum noch Raum für Beteuerungen um die Privatsphäre der Nutzer. Bereits legale Vorgehensweisen, wie z.B. der Einsatz von ‚Cookies', werden vielfach in Frage gestellt: So sammeln *Microsoft* (*bing*), *Google* und *Yahoo* nicht nur Daten der Nutzer, sondern setzen diese durch gespeicherte Profile auch derart ein, dass sie die (kommerziellen wie inhaltlichen) Ergebnisse späterer Suchdurchläufe beeinflussen. Die von Edward Snowden zugänglich gemachten Dokumente führten u.a. zu dem Vorwurf, dass solche und andere Daten auch der *National Security Agency* zugänglich sind (vgl. Greenwald 2013). Meyer kritisierte diese Allianz von kommerziellem Tracking und staatlicher Surveillance: „the story does add to a building realization: that commercial and government surveillance are inseparable" (Meyer 2013). Soltani, Peterson und Barton urteilten über eine solche Möglichkeit des ‚Trittbrettfahrens':

> For years, privacy advocates have raised concerns about the use of commercial tracking tools to identify and target consumers with advertisements. The online ad industry has said its practices are innocuous and benefit consumers by serving them ads that are more likely to be of interest to them. […] The revelation that the NSA is piggybacking on these commercial technologies could shift that debate, handing privacy advocates a new argument for reining in commercial surveillance. (Soltani, Peterson & Barton 2013)

Sieben Unternehmen – darunter *Google*, *Apple*, *Facebook*, *Yahoo* und *Microsoft* – sahen sich diesem Vorwurf ausgesetzt, bestritten anfänglich jede Zusammenarbeit und mussten schließlich unterschiedliche ‚Level' von Kooperationsbereitschaft eingestehen (vgl. Miller 2013; Yadron/Gorman 2013). Ende Oktober 2013 nahm *Google* Stellung zu der Veröffentlichung von Dokumenten, die nahelegten, dass die *NSA* darüber hinaus auch unautorisierten Zugriff auf die Kommunikation von *Google*- und *Yahoo*-Datencentern hatte. David Drummond, *Googles* Chief Legal Officer, kommentierte dies:

> We have long been concerned about the possibility of this kind of snooping, which is why we have continued to extend encryption across more and more Google services and links, especially the links in the slide [...] We are outraged at the lengths to which the government seems to have gone to intercept data from our private fibre networks, and it underscores the need for urgent reform. (Irvine/Squires 2013)

Unabhängig davon, ob diese Unternehmen die Daten der *NSA* freigiebig zur Verfügung gestellt haben: Dass alternative Suchmaschinen wie duckduckgo.com, ixquick.com oder blekko.com derzeit erhöht Anfragen verzeichnen (vgl. Chan 2013, Stallmann/Heid 2013, Gropp 2013)[6], deutet ein flexibles Nutzerverhalten an, das auf aktuelle Ereignisse reagiert. Dabei mag es sich um marginale Veränderungen angesichts der immensen Quantitäten von *Google*-Sucheingaben handeln; dennoch sind gewisse Verschiebungen festzustellen. Die genannten

---

[6] Auf *netzpolitik.org* berichtete Moritz Tremmel bereits im Juli 2012 von einer ansteigenden Nutzung alternativer Suchmaschinen, die er sowohl auf damalige Datenschutzskandale als auch auf die steigende Personalisierung von *Google*-Suchergebnissen zurückführte.

alternativen Anbieter setzen explizit auf Versprechen der Privatsphäre und Diskretion: „die diskreteste Suchmaschine der Welt“ (ixquick) und „Search anonymously. Find instantly.“ (duckduckgo; vgl. auch donttrack.us) heißt es hier. Es ist in diesem Zusammenhang zudem wenig überraschend, dass Microsoft in seiner neuen Internet Explorer-Kampagne unter dem Titel „Your privacy is our priority“ gerade die Nutzerkontrolle über persönliche Daten betont (vgl. dazu Bachmann 2013). Dies ist auch als Reaktion auf die zunehmende Popularisierung des TOR-Projekts zu sehen, das Nutzern unter dem Slogan „Anonymity Online“ eine Rückkehr zum anonymen Surfen verspricht und zugleich eine Spielwiese für illegalen Handel eröffnete.

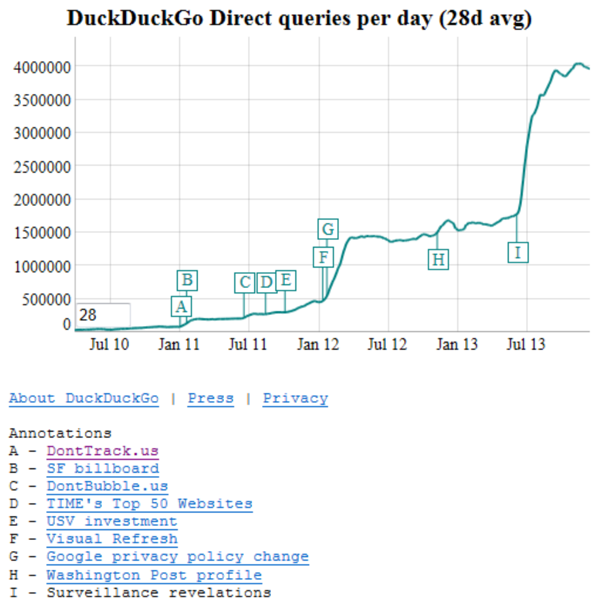


Abb. 5: Visualisierung der *duckduckgo*-Suchanfragen pro Tag (https://duckduckgo.com/traffic.html)

Duckduckgo.com, die als „anti-Google“ (Chan 2013) betitelte Suchmaschine, erhielt im Zuge des Datenskandals Zulauf, was die Vermutung nahelegt, dass diese Eingaben zulasten anderer

Suchmaschinenanbieter zu verzeichnen sind (vgl. ebd.). Auf der Serviceseite selbst findet sich eine Visualisierung der (steigenden) Nutzungshäufigkeiten (siehe Abb. 5).

Es ist zu vermuten, dass sich derartige Änderungen im Nutzerverhalten auch auf die Verlässlichkeit und die Aussagekraft von *Google Trends* auswirken. Anhand der *Trends*-Visualisierungen lässt sich jedoch nicht ablesen, inwiefern Fluktuation und der ‚Verlust' von Nutzern sich auf die Verhältnisse von Suchmaschineneingaben im zeitlichen Verlauf auswirken. Denn der Service enthält keine Angaben darüber, welchen quantitativen Schwankungen die Suchvolumina der *Google-Trends*-Daten unterliegen und wie dies die Ergebnisse, vor allem in Relation zueinander, beeinflusst. Möglich und wahrscheinlich ist, dass die Auswirkungen derzeit quantitativ allenfalls marginal sind; dennoch bleiben sämtliche Einflussfaktoren, die in den Prozess der Datenerhebung fallen, für die (akademische wie kommerzielle) Datenauswertung eine Blackbox. Forschung, die auf *Google Trends* zurückgreift und keinen Zugriff auf die Primärdaten hat, operiert folglich zu einem gewissen Maße blind und kann ausschließlich ergebnisorientierte Aussagen dazu treffen, dass eine Korrelation zwischen spezifischen begrifflichen Sucheingaben besteht, nicht jedoch numerisch genau ermessen, welcher Datenumfang dem zugrunde liegt. Damit ist auch das Analysepotential solcher Arbeiten immens eingeschränkt, insbesondere wenn es zu Metaanalysen der *Google-Trends*-Datengenerierung kommt.

Resümee

Anhand der skizzierten Fälle wurden auf dem ersten Treffen der AG „Daten und Netzwerke" die Potentiale, problematischen Implikationen und soziokulturellen Folgen von suchmaschinengenerierten Big Data diskutiert. Zwar macht *Google Inc.* vermeintlich mehr Zugeständnisse denn je hinsichtlich der Transparenz von Datenspeicherung und -auswertung, doch noch immer sind es nur beschnittene, beschwichtigende Einblicke, die den Nutzern gewährt werden. Die strategisch kommunizierte Öffnung eines Unternehmens, dessen Geschäftsmodell auf der kommerziellen Verwertung transaktionaler Daten beruht, ist nicht zuletzt eine symbolische Geste: In einer pragmatischen Nutzbarmachung täuscht man darüber hinweg, dass nicht die Daten, sondern allenfalls ihre Stellvertreter und Indikatoren offen gelegt werden.

Wissenschaft, die mit *Google Trends* arbeitet, operiert daher gewissermaßen zum Teil ‚blind' – selbst wenn man die Daten an konventionellen quantitativen Erhebungsstandards misst. Zu bedenken ist, inwiefern die Suchmaschinendaten bereits von *Google Inc.* zugerichtet bzw. zensiert werden (etwa durch die Beschränkung einer Häufigkeitsangabe auf eine Skala von 1 bis 100 anstelle exakter numerischer Angaben), und welche Einschränkungen dies für ihre Verwendung in wissenschaftlichen Auswertungen mit sich bringt. Man muss sich überdies fragen, inwiefern bestimmte Themenkomplexe eine prognostische Aussagefähigkeit (un-)wahrscheinlich machen. Für bestimmte Suchanfragen muss eine dominante Signifikanz als historisch-spezifisches Relikt angenommen werden, während andere über eine zeitlich weniger begrenzte Übertragbarkeit verfügen mögen.

Gegen eine umfassende Veröffentlichung von Suchmaschineneingaben spricht zudem aus Sicht des Unternehmens, dass die Publikation bestimmter Korrelationen ggf. die Validität dieser Ergebnisse gefährdet.[7] Denkbar ist auch, dass die von *Google Trends* indizierten ‚angesagten Suchanfragen' sich selbst reproduzieren und verstärken. Dies kann u.a. für kommerzielle Anbieter interessant sein, für die *Google Trends* nicht zuletzt auch eine weitere Onlinewerbeplattform darstellt. Kommerzielle Akteure haben die Bedeutung der Suchmaschinenoptimierung längst erkannt, sodass auch die gezielte Beeinflussung von Suchmaschineneingaben, deren Häufigkeit sowie die Verknüpfung mit anderen Begriffen naheliegend ist. Da der quantitative Umfang von Suchanfragen in *Google Trends* unklar bleibt, kann nicht genau bestimmt werden, welche Größenordnung für solche Eingriffe notwendig wäre. Solche Strategien mögen derzeit unwahrscheinlich anmuten, sie eröffnen jedoch den Blick auf neue softwaretechnische Marktentwicklungen und Anbieter einer ‚virtual mass workforce', wie sie über Plattformen wie *Mechanical Turk* verfügbar sind. Ein prominentes Beispiel für eine – zwar mittlerweile nicht mehr praktikable –

7 Allgemeinere Risiken einer vermeintlich anonymsisierten Veröffentlichung wurden etwa im Fall der 2006 von AOL publizierten Suchanfragen und Nutzerprofile drastisch deutlich (vgl. Arrington 2006).

Manipulationsstrategie stammt etwa von Brent Payne. Der Experte für Suchmaschinenoptimierung brachte 2011 über die Beauftragung dezentraler massenhafter Eingaben von „brent payne manipulated this", via *Mechanical Turk*, dieses Suchergebnis in Googles automatische Vervollständigungsliste (vgl. Vos 2011).

Googles Balanceakt zwischen einer Transparenz von Suchbegriffeingaben sowie den vagen quantitativen Angaben zu Suchvolumina, die kontinuierliche Anpassungen der Algorithmen und die nachdrücklichen Beteuerungen zu Sicherheit und Privatheit der Daten haben somit auch Manipulationsrisiken und Abweichungen im Nutzungsverhalten und somit in den analytischen Potentialen im Blick. So unterschiedlich diese Aspekte des Unternehmens auch scheinen, sie sind gleichermaßen Teil einer Daten-kommerziell ausgerichteten Strategie, um *Googles* Big Data ‚natürlich' zu halten.

Literatur

Arrington, Michael (2006): „AOL Proudly Releases Massive Amounts of Private Data“, in: *Techcrunch*, http://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data, publ. 06.08.2006, zit. 18.12.2013.

Bachmann, Katy (2013): „New Microsoft Privacy Campaign Promotes Consumer Control. Campaign will stir debate over Do Not Track“, in: *Adweek*, http://www.adweek.com/news/advertising-branding/new-microsoft-privacy-campaign-promotes-consumer-control-148781, publ. 22.04.2013, zit. 11.09.2013.

Butler, Declan (2013): „When Google got flu wrong“, in: *Nature. International Weekly Journal of Science* 4944. 7436, www.nature.com/news/when-google-got-flu-wrong-1.12413, publ. 13.02.2013, zit. 24.06.2013.

Chan, Xenia (2013): „The anti-Google: DuckDuckGo search engine gets traffic boost after Snowden leaks“, in: *South China Morning Post: Technology*, http://www.scmp.com/lifestyle/technology/article/1264971/anti-google-duckduckgo-search-engine-gets-traffic-boost-after, publ. 18.06.2013, zit. 24.06.2013.

Choi, Hyunyoung/Hal Varian (2011): *Predicting the Present with Google Trends*, http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf, publ. 18.12.2011, zit. 24.06.2013.

Cook, Samantha u.a. (2011): „Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic“, in: *Plos One*, www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0023610, publ. 19.08.2011, zit. 24.06.2013.

DuVander, Adam (2012): „Google Trends API: Developers Want It, Says Google Insights“, in: *ProgrammableWeb*, , http://blog.programmableweb.com/2012/02/10/google-trends-api-developers-want-it-says-google-insights/#ixzz2lInhfDmo, publ. 10.02.2012, zit. 21.11.2013.

Eysenbach, Gunther (2002): „Infodemiology: The Epidemiology of (Mis)information“, in: *American Journal of Medicine* 113, S. 763-765.

– (2006): „Infodemiology: Tracking Flu-Related Searches on the Web for Syndromic Surveillance“, in: *AMIA Annual Symposium, Proceedings*, S. 244-248.

– (2009): „Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet“, in: *Journal of Medical Internet Research* 11.1, http://www.jmir.org/2009/1/e11/, zit. 14.10.2013.

Google Flu Trends (2011): *Explore flu trends around the world*, http://www.google.org/flutrends/, zit. 18.12.2013.

Google Trends Support (2013): *So werden Trends-Daten siert*, https://support.google.com/trends/answer/87284?hl=de&ref_topic=13975, zit. 18.12.2013.

Greenwald, Glenn (2013): „NSA Prism program taps in to user data of Apple, Google and others“, in: *The Guardian*, www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data, publ. 07.06.2013, zit. 24.06.2013.

Gropp, Martin (2013): „Alternative Internetunternehmen: Profiteure der Enthüllung“, in: *Frankfurter Allgemeine tung*, http://www.faz.net/aktuell/wirtschaft/unternehmen/alternative-internetunternehmen-profiteure-der-enthuellungen-12280219.html, publ. 13.07.2013, zit. 11.09.2013.

Heisler, Kevin (2008): „Google Trends Goes Daily; Drops Hourly Updates“, in: *Search Engine Watch,* http://searchenginewatch.com/article/2054079/Google-Trends-Goes-Daily-Drops-Hourly-Updates, publ. 05.06.2008, zit. 18.12.2013.

Hwang, Heej (2008): „A new flavor of Google Trends”, in: *Google Official Blog,* http://googleblog.blogspot.nl/2008/06/new-flavor-of-google-trends.html, publ. 10.06.2008, zit. 18.12.2013.

Irvine, Chris/Nick Squires (2013): „NSA ‚tapped into Google and Yahoo data centres‘“, in: *The graph*, www.telegraph.co.uk/news/worldnews/northamerica/usa/10416025/NSA-tapped-into-Google-and-Yahoo-data-centres.html, publ. 30.10.2013, zit. 18.12.2013.

Leinweber, David (2013): „Big Data gets Bigger“, in: *Forbes Magazine*, www.forbes.com/sites/davidleinweber/2013/04/26/big-data-gets-bigger-now-google-trends-can-predict-the-market, publ. 26.04.2013, zit. 24.06.2013.

Manovich, Lev (2011/2012): „Trending: The Promises and the Challenges of Big Social Data“, http://www.manovich.net/DOCS/Manovich_trending_paper.pdf, publ. 23.04.2011, zit. 13.09.2013; nicht mehr online. Inzwischen publ. in: Matthew K. Gold (Hg.): *Debates in the Digital Humanities*, Minneapolis/London 2012, S. 460-475.

Mayer, Marissa (2006): „Yes, we are still all about search”, in: *Google Official Blog,* http://googleblog.blogspot.de/2006/05/yes-we-are-still-all-about-search.html, publ. 10.05.2006, zit.18.12.2013.

Meyer, David (2013): *NSA's taste for cookies reveals the danger of marketing-driven web tracking*, http://gigaom.com/2013/12/11/nsas-taste-for-cookies-reveals-the-danger-of-marketing-driven-web-tracking/, publ. 11.12.2013, zit. 10.06.2014.

Miller, Claire Cain (2013): „Tech Companies Concede to Surveillance Program”, in: *The New York Times*, http://www.nytimes.com/2013/06/08/technology/tech-companies-bristling-concede-to-government-surveillance-efforts.html?pagewanted=all, publ. 07.06.2013, zit. 02.12.2013.

Millis, Elinor (2007): „Google Trends API coming soon Google to release API for Google Trends and let you download data from the program into a spreadsheet”, in: *CNET,* http://news.cnet.com/8301-10784_3-9828916-7.html, publ. 04.12.2007, zit. 21.11.2013.

Nature editorial board (2007): „A matter of trust: Social scientists studying electronic interactions must“, in: *Nature* 449. 7163, S. 637-638, http://www.nature.com/nature/journal/v449/n7163/pdf/449637b.pdf, zit. 10.09.2014.

Polgreen, Philip M. u.a. (2008): „Using Internet Searches for Influenza Surveillance“, in: *Clinical Infectious Diseases*, Vol. 47(11), S. 1443-1448.

Preis, Tobias/Hellen S. Moat/H. Eugene Stanley (2013): „Quantifying Trading Behavior in Financial Markets Using Google Trends“, in: *Scien-*

tific Reports 3.1684, www.nature.com/srep/2013/130425/srep01684/pdf/srep01684.pdf, zit. 24.06.2013.

Preis, Tobias/Daniel Reith /Eugene Stanley (2010): „Complex dynamics of our economic life on different scales", in: *Philosophical Transactions A* 368, S. 5707-5719, http://rsta.royalsocietypublishing.org/content/368/1933/5707.full, publ. 15.11.2010, zit. 24.06.2013.

Schnitt, Barry (2001): *2001 Year-End Google Zeitgeist: Search patterns, trends, and surprises*, http://www.google.de/press/zeitgeist2001.html, zit. 18.12.2013.

Soltani, Ashkan/Andrea Peterson/Barton Gellman (2013): „NSA uses Google cookies to pinpoint targets for hacking" in: *The Washington Post*,, http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/nsa-uses-google-cookies-to-pinpoint-targets-for-hacking, publ. 10.12.2013, zit. 18.12.2013.

Stallmann, Stephanie E./Tatjana Heid (2013): „NSA-Skandal beeinflusst Surfverhalten. Mega-Run auf Alternativen zu Google in Deutschland", in: *Focus Online*,, www.focus.de/digital/internet/nsa-skandal-schadet-google-mega-run-auf-alternative-suchmaschinen-in-deutschland_aid_1063315.html, publ. 05.08.2013, zit. 11.09.2013.

Tremmel, Moritz (2012): „Alternative Suchmaschinen legen zu", in: *netzpolitik.org*, https://netzpolitik.org/2012/alternative-suchmaschinen-legen-zu, publ. 03.07.2012, zit. 24.06.2013.

Vos, Cayley (2011): „Volume of Queries for a Term (4.63)", in: *Beat the Autocomplete,* http://www.beattheautocomplete.com/study-results/1-volume-of-queries-for-a-term-4-63.html, publ. 22.02.2011, zit. 10.06.2014.

Yadron, Danny/Siobhan Gorman (2013): „Tech Firms Race to Encrypt More Data After NSA Leak", in: *The Wall Street nal*, http://online.wsj.com/news/articles/SB10001424052702304073204579170080794610094, publ. 31.10.2013, zit. 02.12.2013.