Abschlussbericht zum Forschungsprojekt **#Datenspende: Google und die Bundestagswahl 2017**

Wer sieht was? Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine

Autoren: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig

Kaiserslautern, Juni 2018

Projektpartner: Bayerische Landeszentrale für neue Medien (BLM), Medienanstalt Berlin-Brandenburg (mabb), Hessische Landesanstalt für privaten Rundfunk und neue Medien (LPR Hessen), Landeszentrale für Medien und Kommunikation Rheinland-Pfalz (LMK), Landesmedienanstalt Saarland (LMS), Sächsische Landesanstalt für privaten Rundfunk und neue Medien (SLM), Algorithm Accountability Lab der TU Kaiserslautern, AlgorithmWatch gGmbH

Editorial

Algorithmen lenken zunehmend unsere Aufmerksamkeit. Sie kennen unsere Vorlieben und beeinflussen unsere Medienwahrnehmung. Daher müssen wir uns mit der Macht der Algorithmen – gerade in Bezug auf die politische Meinungsbildung in unserer demokratischen Gesellschaft – mehr denn je auseinandersetzen.

Wie relevant soziale Netzwerke und Suchmaschinen als Agendasetter mittlerweile sind, haben gerade wieder die im April veröffentlichten Ergebnisse der MedienGewichtungsStudie der Medienanstalten belegt. Demnach entfällt mehr als die Hälfte ihrer Tagesreichweite auf die meinungsbildungsrelevante Nutzung. Vor allem die beiden Plattformen Facebook und Google können die öffentliche Meinung bereits erheblich steuern.

Die Frage der Personalisierung stand im Fokus der Studie „#Datenspende: Google und die Bundestagswahl 2017“, deren Abschlussbericht vor Ihnen liegt. Die Informatik-Professorin Dr. Katharina Zweig von der Technischen Universität Kaiserslautern untersuchte dafür den Einfluss des Google-News-Algorithmus auf die Bundestagswahl im vergangenen Jahr.

Vorab will ich an dieser Stelle nur zwei Punkte kurz anreißen: Überraschend war, dass die Google-Suche nach Politikern allen Nutzern annähernd die gleichen Ergebnisse liefert. Das Forschungsprojekt misst nur wenige Anzeichen für Personalisierung. Klar ist aber auch: Die Studie war eine Momentaufnahme. Bereits morgen können die Ergebnisse anders aussehen.

Noch entscheidender als die Ergebnisse dieser Momentaufnahme ist aber: Das Setup der Studie zeigt, wie man Kontrolle leisten kann – ohne dass Anbieter rechtlich dazu verpflichtet werden, Algorithmen komplett transparent zu machen. Auch dazu finden Sie im Folgenden Details.

Was bedeutet das alles nun für Aufsicht und Regulierung? Gesellschaftlicher Konsens ist mittlerweile: Eine Demokratie muss sich Transparenz darüber verschaffen können, in welchem Ausmaß Suchmaschinen oder andere Intermediäre meinungsrelevante Auswahlentscheidungen vorwegnehmen. Nur so können wir gegensteuern, etwa wenn Algorithmen die Verbreitung von Falschmeldungen oder Hetze fördern.

Deshalb begrüßen es die Medienanstalten sehr, dass Bestimmungen zur Regulierung von Informationsintermediären in den 22. Rundfunkänderungsstaatsvertrag aufgenommen werden sollen. Sie sind bereit, in dem Bereich Verantwortung zu übernehmen. Im Mai dieses Jahres haben wir deshalb Transparenz, Diskriminierungsfreiheit und Berichtspflicht als Mindeststandards für die Regulierung von Informationsintermediären formuliert.

Abschließend noch ein Wort des Dankes: Das Forschungsprojekt „#Datenspende“ wurde von den Landesmedienanstalten Bayern (BLM), Berlin-Brandenburg (mabb), Hessen (LPR Hessen), Rheinland-Pfalz (LMK), Saarland (LMS) und Sachsen (SLM) finanziert und in Kooperation mit der gemeinnützigen Initiative für die Kontrolle algorithmischer Entscheidungsfindung AlgorithmWatch sowie der TU Kaiserslautern durchgeführt. Ohne das große Engagement aller Beteiligten – und ganz besonders Frau Prof. Zweig – wäre die Studie nicht möglich gewesen.

München, im Juni 2018
Siegfried Schneider
Präsident der Bayerischen Landeszentrale für neue Medien (BLM)

Executive Summary

Der vorliegende Bericht analysiert die Google-Suchergebnisse von mehr als 1.500 freiwilligen Datenspendern, die in den fünf Wochen vor der Bundestagswahl am 24. September 2017 alle vier Stunden automatisiert mit Google nach 16 vordefinierten Namen von Parteien und Politikern gesucht haben. Er basiert auf einem bereinigten Datenbestand von über 8.000.000 Datensätzen, die im Rahmen des Forschungsprojekts „#Datenspende: Google und die Bundestagswahl 2017“ erzeugt und uns zur Auswertung übermittelt wurden. Das #Datenspende-Projekt wurde im Auftrag von sechs Landesmedienanstalten durchgeführt. Medienpartner war Spiegel Online. Schwerpunkte der vorliegenden Untersuchung sind u. a. die Frage nach dem Personalisierungsgrad der Suchergebnisse, dem Anteil der Regionalisierung sowie nach der Gefahr einer algorithmenbasierten Filterblasenbildung oder –verhärtung durch den Marktführer im Suchmaschinenmarkt.

Hauptergebnisse

Die erste Ergebnisseite bei der Suche nach Nachrichten mit Google News und die von Google im Rahmen der organischen Suche ausgerollten Schlagzeilen werden insgesamt von etablierten Medienhäusern, insbesondere aus dem Printbereich, dominiert. Ausnahmen stellen vor allem die Online-Zeitung Huffington Post sowie das Freemail Portal t-online.de dar, die häufig als Nachrichtenquellen auftauchen.

Bei der Google-Suche nach Parteien führt die Mehrheit der Suchergebnisse auf der ersten Ergebnisseite auf Webseiten, deren Inhalte die Parteien selbst kontrollieren. Bei 34 % der Treffer handelt es sich unmittelbar um Webseiten von Parteien, Parteiangehörigen oder Ortsverbänden. Weitere 17 % der Suchergebnisse sind Social-Media-Profile der Parteien, deren Inhalte ebenfalls direkt kontrolliert werden können. Auf Wikipedia-Einträge, die zumindest mitbestimmbar sind, verlinken 11 % der Treffer. Medienangebote machen demgegenüber 26 % der Ergebnisse bei der Google Suche nach Parteien aus (vgl. Abschnitt 3.4).

Die Parteien sind sehr unterschiedlich gut darin, selbst-kontrollierte Webseiten wie eigene Homepages oder Social-Media-Accounts auf die erste Suchergebnisseite zu bringen. Während beim Suchbegriff „Bündnis90/Die Grünen“ insgesamt über 80 % der Treffer auf parteieigene Websites, Social-Media-Profile oder den Wikipedia-Eintrag der Partei führen, sind es bei der Suche nach AfD nur knapp 27 %. Bei der Suche nach „Die Linke“ liegt der Anteil selbst-kontrollierter Treffer bei rund 82 %, bei der FDP bei 75 %, bei der CDU sind es 63 %, bei der SPD 52 % und bei der CSU rund 52 % (vgl. Abschnitt 3.4).

Mit großem Abstand (80 %) werden die Onlineableger klassischer deutscher Printmedien (Tageszeitungen und Zeitschriften) am Häufigsten von Google in den Schlagzeilen ausgerollt, gefolgt von solchen Medienangeboten, deren einziger Verbreitungsweg das Internet ist, mit 12 %. Inhalte aus dem öffentlich-rechtlichen Rundfunk machen 4 % der Schlagzeilen bei Google aus (vgl. Abschnitt 3.4).

Die Google-Suchergebnislisten zu prominenten Politikern weisen im Wesentlichen sehr hohe Ähnlichkeit auf. Im Durchschnitt unterscheiden sich nur 1 bis 2 der Ergebnisse von im Mittel 9 der sogenannten organischen Suchergebnisse bei der Suche nach dem selben Suchbegriff. Die Suchergebnisse von Parteien sind zunächst etwas weniger ähnlich. Hier unterscheiden sich im Durchschnitt 3-4 der organischen Suchergebnisse (vgl. Abschnitt 4.2). Berücksichtigt man allerdings die stärkere Regionalisierung der Suchergebnisse, z. B. durch Webseiten von Ortsverbänden der Parteien, bleiben auch hier durchschnittlich 1-2 unterschiedliche organische Suchergebnisse übrig (vgl. Abschnitt 5).

Bei Google News ist der durchschnittliche Anteil nicht geteilter Suchergebnisse ähnlich niedrig (vgl. Abschnitt 4.3). Es gibt in den Suchergebnissen insgesamt also wenig Raum für Personalisierung. Die Grundlage für eine algorithmenbasierte Bildung oder Verhärtung von Filterblasen im Sinne der Theorie von Eli Pariser ist damit auf Basis der vorliegenden Untersuchung auf Google nicht gegeben.

Bei der Suche nach einigen Parteien und Politikern, wie bspw. SPD oder Angela Merkel, zeigt Google so gut wie allen Nutzern zu allen Suchzeitpunkten als oberste Ergebnisse Schlagzeilen an. Demgegenüber werden für die Suchbegriffe „Dietmar Bartsch“ und „Katrin Göring-Eckardt“ zu manchen Zeitpunkten an so gut wie alle Nutzer Ergebnislisten ohne Schlagzeilen ausgerollt (vgl. Abschnitt 6.1).

Es gibt immer wieder kleine Gruppen von Personen mit stärker abweichenden Suchergebnislisten. Die wahrscheinlichste Erklärung dafür findet sich in einer vom Nutzer im Google-Account vorgenommenen Einstellung einer anderen Suchsprache, z. B. Englisch oder Französisch. Somit handelt es sich bei diesen Gruppen mit stark abweichenden Ergebnislisten weniger um einen Personalisierungseffekt als um das Ergebnis eines Sprachfilters (vgl. Abschnitt 7).

Die von Google ausgelieferten Ergebnislisten sind im Wesentlichen bestimmt von der Suchspracheneinstellung, dem Standort (z.B. durch IP-Adresse) und der allgemeinen Relevanz der Webseiten (vgl. Abschnitt 7).

Das #Datenspende-Projekt zeigt, dass die Gesellschaft einen wichtigen, Öffentlichkeit schaffenden und die Meinungsbildung potenziell beeinflussenden, Algorithmus gemeinsam auf relevante Phänomene untersuchen kann, ohne dafür Einblick in den dahinterliegenden Code zu benötigen. Notwendig dafür ist jedoch die Schaffung geeigneter Schnittstellen zur Überprüfung der algorithmischen Selektionsleistungen von sozialen Netzwerken und anderen Intermediären (vgl. Abschnitt 8).

Inhaltsverzeichnis

1 Einleitung: Gefahren der Filterblase für die Gesellschaft 6
1.1 Das Modell der algorithmisch erzeugten und verstärkten Filterblasen 6
1.1.1 Filtern von Suchergebnissen mit Personalisierung und Regionalisierung 6
1.1.2 Wann sind algorithmisch erzeugte und verstärkte Filterblasen gefährlich? 7
1.2 Überprüfung der Filterblasen-Theorie von Eli Pariser 8
2 Studiendesign 10
2.1 Aufbau der Software und Enrollment 10
2.2 Struktur der Daten und wichtige Begriffe 12
2.3 Datenaufbereitung 14
3 Überblick über die Daten 15
3.1 Zeitliche Verteilung der Suchanfragen 15
3.2 Geographische Verteilung der Datenspenderinnen und -spender 18
3.3 Verteilung von Internet- und Medienangeboten in den Suchergebnissen 19
3.4 Owned Content, Social Media und Medienangebote 25
4 Raum für Personalisierung / Regionalisierung 29
4.1 Anteile gleicher Suchergebnisse 29
4.2 Durchschnittliche Anzahl gemeinsamer Links 31
4.3 Raum für Personalisierung bei Google News 34
5 Regionalisierung 35
6 Dynamik 36
6.1 Dynamik des Vorkommens von Schlagzeilen 36
6.2 Dynamik der Anzahl unterschiedlicher Top-Level-Domains 39
6.3 Dynamik der Anteile editierbarer Beiträge 42
7 Suchergebnislisten mit geringer Ähnlichkeit zu allen anderen 44
7.1 Beispiel: Suchergebnisse von einer französischen IP aus 45
7.2 Beispiel 2: Suche aus Deutschland mit partiell russischen Suchergebnissen 46
7.3 Erklärungsansatz 52
8 Zusammenfassung 53
8.1 Generalisierung der Ergebnisse: Möglichkeiten und Grenzen 55
8.2 Forderung nach geeigneten Schnittstellen zur Überprüfung der Filterblasentheorie in sozialen Netzwerken und auf anderen Intermediären 55
9 Danksagung 56
10 Quellen 56
Appendix 57
Hauptkategorien 57
Unterkategorien 58

1 Einleitung: Gefahren der Filterblase für die Gesellschaft

Die politische Meinungsbildung sowie der generelle Zugang zu Informationen haben sich durch die Digitalisierung stark verändert. Informationsquellen wie Zeitung, TV und Radio, bei denen große Teile der Bevölkerung dieselben Nachrichten und Interpretationen dieser Nachrichten lasen, hörten oder sahen, werden im Zuge des digitalen Wandels durch immer vielfältigere und personalisiertere Medienangebote abgelöst. Dabei wird die Personalisierung durch algorithmenbasierte Systeme ermöglicht: So entscheidet z.B. ein Algorithmus – und kein Mensch – darüber, welche Beiträge Nutzer interessieren könnten, und nur diese werden ihnen in sozialen Netzwerken angezeigt. Dasselbe gilt auch für personalisierte Suchmaschinen wie Google, Yahoo oder Bing, für welche es unmöglich ist, bei über 3.200 Milliarden Suchanfragen im Jahr 2016 (Statista, 2016) eine menschengemachte Sortierung vorrätig zu halten.

1.1 Das Modell der algorithmisch erzeugten und verstärkten Filterblasen

Die Möglichkeiten und Gefahren einer sogenannten algorithmisch erzeugten Filterblase steigen damit stetig. Unter diesem Begriff ist ein Teilkonzept der Filterblasentheorie von Eli Pariser zu verstehen. In seinem 2011 veröffentlichten Buch „The Filter Bubble: What the Internet Is Hiding from You“ (Pariser, 2011) wies der Internetaktivist auf die möglichen Gefahren sogenannter Filterblasen hin. Er zeigte in seinem TED-Talk anhand zweier Screen-Shots von 2011, dass zwei seiner Freunde deutlich unterschiedliche Ergebnisse bei der Suche nach dem Begriff „Egypt“ auf der Onlinesuchplattform Google erhalten hatten[1]. Daraus entwickelte er eine Theorie, wonach personalisierte Algorithmen in den sozialen Medien dazu neigen, Individuen Inhalte anzuzeigen, die mit den bisherigen Ansichten des Benutzers übereinstimmen, sodass sich verschiedene Informations-Sphären bilden können, in denen unterschiedliche Inhalte oder Meinungen vorherrschen.

Kurz gesagt, kann eine individuelle Filterung des Informationsflusses dazu führen, dass Gruppen oder Individuen unterschiedliche Fakten präsentiert bekommen, also in einem einzigartigen Informationsuniversium leben[2]. Dies ist insbesondere dann besorgniserregend, wenn die Inhalte politisch extremer Natur sind und durch die einseitige Perspektive die Diskursfähigkeit der Bürger geschwächt wird oder gänzlich verkommt. Eine **Filterblase** ist in diesem Sinne eine Selektion von Nachrichten, die den eigenen Perspektiven entsprechen, was insbesondere im politischen Bereich zu einer Verhärtung der eigenen politischen Position führen könnte[3].

1.1.1 Filtern von Suchergebnissen mit Personalisierung und Regionalisierung

Da die Anzahl der prinzipiell mit einem Suchbegriff inhaltlich verbundenen Webseiten für die meisten Anfragen mehr als 10 beträgt und gleichzeitig die ersten 10 gezeigten Webseiten die größte Aufmerksamkeit durch die Nutzer erfahren, ist es unerlässlich, dass Suchmaschinen die möglichen Suchergebnisse filtern. Einen der wichtigsten **Filter** stellt dabei sicherlich die Sprache des Nutzers dar, daneben spielen Aktualität und Popularität der Webseite eine Rolle, vermindert auch ihre Einbettung in das gesamte WWW (gemessen z. B. durch den PageRank).

---

[1] Zu sehen in Eli Parisers TED-Talk „Beware the „filter bubbles", https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles (Abgerufen am 12.5.2018).

[2] “a unique universe of information for each of us” (Pariser, 2011, S. 9).

[3] Unter „Filterblasen“ kann man auch sehr viel weitergehende Konzepte verstehen. Die Selektion von Webseiten durch staatliche Zensur kann ebenfalls durch Informationsverengung eine Filterblase erzeugen. Diese Zensur wird zwar auch durch Algorithmen unterstützt, aber es handelt sich nicht per se um eine durch algorithmenbasierte Personalisierung erzeugte Filterblase, von der wir in diesem Text ausgehen. Diese Art der Informationseinengung und ihre möglichen Folgen für eine Filterblasenbildung werden hier nicht untersucht. Nicht zuletzt könnte ein Suchmaschinenbetreiber oder eine soziale Medien-Plattform ganz bewusst die Datenbasis in einer Richtung einschränken und damit zwar allen Nutzern dasselbe präsentieren, aber nur einen selektiven Ausschnitt aus der Wirklichkeit darbieten. Diese Möglichkeit untersuchen wir hier nicht.

Ein im Rahmen von Eli Parisers Filterblasentheorie besonders wichtiger Mechanismus der Filterung ist die „Personalisierung“. Dabei folgen wir beim Begriff der (vorselektierten) **Personalisierung** den Ausführungen von Zuiderveen Borgesius et at. (2016), wonach eine Personalisierung das Auswählen von Inhalten ermöglicht, die bisher noch nicht von der Nutzerin oder dem Nutzer angeklickt wurden, deren Inhalte aber Personen mit ihren oder seinen Interessen zugeordnet werden. Dahinter stecken algorithmisch gesehen sogenannte "Empfehlungssysteme" (recommendation systems), die z.B. aus dem Klickverhalten anderer Personen, die in der Vergangenheit ein ähnliches Klickverhalten wie die momentan suchende Person zeigten, schlussfolgern, was diese interessieren könnte. Denkbar ist auch, dass für jede Person entsprechend ihres eigenen Klickverhaltens zusammen mit bekannten Kategorisierungen der von ihr angeklickten Inhalte ein Profil zusammengestellt wird, das beispielsweise besagt: "Diese Person bevorzugt Nachrichten über Sport und Wirtschaft, liest eher mittellange Texte und Nachrichten, die nicht älter sind als einen Tag." (Weare, 2009). Beide Wege können bei den enormen Nutzerzahlen nur algorithmisch über verschiedene Verfahren des maschinellen Lernens erzielt werden und daher nur statistische Modelle bilden (Zweig et al., 2017). Es ist zu erwarten, dass Nutzer, die in ihre Google-Konten eingeloggt sind, tendenziell stärker personalisierte oder überhaupt personalisierte Suchergebnisse bekommen.

Von der Personalisierung abgegrenzt sind solche Webseiten in der Suchergebnisliste, die ein Nutzer oder eine Nutzerin zu einem früheren Zeitpunkt für dieselbe Suche schon einmal angeklickt hat. Für Außenstehende sind diese zwar nicht von den durch Algorithmen personalisierten Einträgen zu unterscheiden, inhaltlich tragen sie aber nicht zu algorithmenbasierten Filterblasenbildungen oder -verstärkungen bei, da es sich um Inhalte handelt, für die sich die Nutzer früher schon einmal entschieden haben.

Als **Regionalisierung** bezeichnen wir die Auswahl von Webseiten für eine ganze Gruppe von Personen, die momentan von einer bestimmten Region aus suchen oder bekanntermaßen aus einer bestimmten Region stammen, die aber nicht notwendigerweise eine Region in ihrer Suchanfrage nennen. Der momentane Standort kann z. B. grob abgeleitet werden aus der IP-Adresse des Gerätes, von dem aus gesucht wird, aber auch genauer aus Standortinformationen bei Smartphones stammen oder aus dem der Suchmaschine bekannten Profil (Teevan et al., 2011). Die ausgelieferten Webseiten selbst haben dabei einen klaren Bezug zum von Google bestimmten Ort des Interesses; das ist beispielsweise dann der Fall, wenn der Name eines nahegelegenen Ortes mehrfach auf der Webseite auftaucht.

Es ist wichtig zu bemerken, dass eine sehr kleinteilige Regionalisierung zu einer Personalisierung gerechnet werden kann – wenn z. B. an alle Personen eines Haushaltes jeweils eine Auswahl regionaler Webseiten ausgeliefert wird, die sich jeweils von der Auswahl für ihre Nachbarn unterscheidet.

1.1.2 Wann sind algorithmisch erzeugte und verstärkte Filterblasen gefährlich?

Eli Parisers Filterblasen-Theorie mit ihren beunruhigenden Folgen für die Gesellschaft basiert auf diesen vier grundlegenden Mechanismen**:**

1. **Personalisierung:** Die auf eine Person zugeschnittene Selektion von Inhalten, die durch Algorithmen eine neue Granularität und vorher unbekannte Skalierbarkeit erhalten.
2. **Geringe Überlappung der jeweiligen Nachrichtenresultate**: Eine geringe oder gar nicht vorhandene Überlappung der Filterblasen, sodass Nachrichten und Informationen der einen Gruppe in einer anderen unbekannt bleiben.
3. **Inhalte:** Die Natur dieser Inhalte, die im Wesentlichen erst bei politisch brisanten Themen und bei hochgradig unterschiedlichen Perspektiven problematisch werden.
4. **Isolation von anderen Informationsquellen:** Die Personengruppen, deren Nachrichtenlage jeweils homogene, politisch brisante und einseitige Perspektiven aufweist, nutzen kaum andere Informationsquellen oder nur solche, die *sie in extrem ähnliche Filterblasen stecken.*

Je stärker diese vier Mechanismen ausgeprägt sind, desto stärker wird der Filterblaseneffekt, einschließlich seiner schädlichen Folgen für die Gesellschaft. Von entscheidender Bedeutung ist allerdings der Personalisierungsgrad, denn bei geringer Personalisierung eines nachrichtenselektierenden Algorithmus kommt es gar nicht erst zu politisch relevanten Filterblasen. Wenn der Personalisierungsgrad höher oder gar hoch ist und Filterblasen nachweisbar sind, müssen sie jedoch nicht zwingend politisch wirksam werden, wenn entweder die Inhalte nicht politischer Natur sind oder wenn Nutzer auch andere Informationsquellen nutzen. So sind beispielsweise Informationen, die Bürger unterschiedlicher Sprache erhalten, definitionsgemäß in den meisten Fällen überlappungsfrei, wenn die Ergebnisse in eben jenen Sprachen sind, trotzdem stecken diese Bürger inhaltlich gesehen noch lange nicht in einer Filterblase.

1.2 Überprüfung der Filterblasen-Theorie von Eli Pariser

Da Algorithmen in der Lage sind, den Informationsfluss an Nutzerinnen und Nutzer zu steuern, kommt ihnen eine ähnliche Torwächter-Funktion („gatekeeper“) zu (Vīķe-Freiberga et at., 2013), wie im klassischen Journalismus den Journalistinnen und Journalisten (siehe hierzu die Ausführungen von Moe & Syvertsen, 2007). Daher ist es notwendig zu überprüfen, ob und wenn ja, wie stark die von Algorithmen auf den verschiedenen Intermediären und Suchmaschinen erzeugten und verhärteten Filterblasen sind. Die Anzahl der belastbaren Studien ist aber eher gering: Eine für Deutschland besonders wichtige Studie vom Hans-Bredow-Institut beantwortet die Frage nach dem Informationsmix positiv: Zum heutigen Zeitpunkt seien die Informationsquellen divers und können daher andere Nachrichten und Informationen einer eventuell algorithmisch erzeugten oder verhärteten Filterblase durchdringen (Schmidt et al., 2017). Hierbei wird sogar angeführt, dass Algorithmen eine Möglichkeit darstellen, Filterblasen aufzusprengen (Bozdag & van den Hoven, 2015), sollte eine solche Funktionalität explizit in ihnen verankert werden.

Eine quantitative Auswertung des Personalisierungsgrades für eine größere Nutzerbasis wurde bis 2017 unseres Wissens nach nicht durchgeführt, nur anekdotische Überprüfungen: So bat beispielsweise Jacob Weisberg für einen Slate-Artikel fünf Personen nach Themen zu suchen und stellte fest, dass die Resultate sehr ähnlich waren (Weisberg, 2011).

Erst mit einer großen Nutzerbasis können aber die wichtigen Fragen nach dem Personalisierungsgrad und der Überlappung der einzelnen Nachrichtenströme geklärt werden. Die Durchführung einer derartigen Untersuchung auf Facebook wäre angesichts der Debatte, die 2016 nach dem Wahlsieg Donald Trumps zum US-amerikanischen Präsidenten über den Einfluss von Filterblasen in sozialen Netzwerken angefacht wurde, dringend notwendig, ist aber aufgrund mangelnder APIs[4] momentan nicht möglich.
Anlässlich des politischen Großereignisses der Bundestagswahl 2017 entschlossen wir uns zur Durchführung des Projekts *#Datenspende: Google und die Bundestagswahl 2017*, um herauszufinden, ob Google die Suchergebnisse bereits personalisiert, wie vielfach spekuliert wird. Zudem ist es deutlich einfacher, eine solche Untersuchung auf Suchmaschinen zu führen, deren Ergebnis auf einer HTML-Seite vorliegt, die leicht weiterverarbeitet werden kann.

Mit dem Projekt haben wir zum einen ein Studiendesign vorgestellt, das diese Frage automatisiert für eine beliebig große Stichprobe von Nutzerinnen und Nutzern von Suchmaschinen und beliebige Suchanfragen beantworten kann. Durch das Design konnten die Nutzerinnen und Nutzer jederzeit nachvollziehen, welche Suchanfragen wir über ihren Account an Google stellten. Damit war die Datensammlung vertrauenswürdig und mehr als 4000-mal wurde die dafür notwendige Software heruntergeladen.

Weiterhin stellt das Studiendesign somit einen proof-of-concept dar, dass es der Gesellschaft grundsätzlich möglich ist, dauerhaft den Personalisierungsgrad von Suchmaschinen für beliebige Suchbegriffe zu überwachen. Das generelle Design kann auch auf andere Intermediäre übertragen werden, wenn

[4] Unter einer API versteht man eine Schnittstelle, mit der automatisiert mit einem Programm interagiert werden kann. So könnten z.B. mittels einer API Informationen ohne Umwege aus Datenbanken abgefragt werden.

entsprechende APIs den selektiven Zugriff auf die für eine Studie relevanten Inhalte beschränkt, um eine ähnliche Vertrauenswürdigkeit aufzubauen. Auf Facebook wäre dies z. B. ein selektiver Zugriff auf die Nachrichten von Medien in einer Timeline oder die politische Wahlwerbung, die den Zugriff auf die privaten Nachrichten von Freunden ausschließt.

2 Studiendesign

Das Studiendesign sowie die Grundlagen der Erhebung werden im Folgenden ausgeführt, hierzu gehören auch die Struktur der Daten, wichtige Begrifflichkeiten und die Aufbereitung der Datenbasis.

2.1 Aufbau der Software und Enrollment

Das Basiswerkzeug, mit dem die Daten gesammelt wurden, ist ein Plug-In, das leicht in Internetbrowser integriert werden kann und dann den Browser des spendenden Nutzers verwendet, um automatisiert

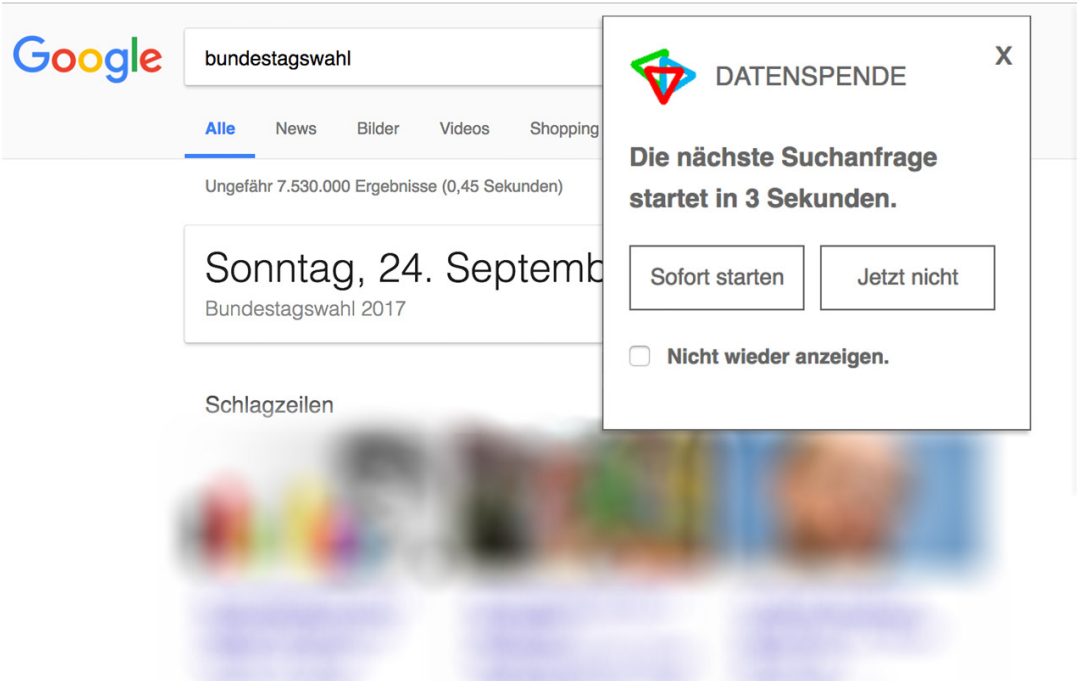


Abbildung 1: Browser-Plug-In, kurz bevor eine Suche bei einem Nutzer ausgelöst wurde, dessen erste Suchergebnisseite dann an die bereitgestellte Serverstruktur weitergeleitet und damit „gespendet" wurde.

Suchanfragen zu stellen und die Daten an einen zentralen Server zu schicken. Das gemeinsam mit der Firma *‚lokaler'* erstellte Plug-In wurde für die Internetbrowser Chrome und Firefox bereitgestellt, um eine Marktabdeckung in Deutschland von über 60 % zu erreichen (Statista, 2018). Alle notwendigen Einblicke in den Quellcode des Plug-Ins wurden bei Projektbeginn veröffentlicht[5]. Das Plug-In suchte zu festen Suchzeitpunkten (4:00, 8:00, 12:00, 16:00, 20:00 und 24:00 Uhr) nach 16 Suchbegriffen, wenn der Browser zu diesem Zeitpunkt geöffnet war[6]. Die Suchanfragen an Google und Google News verliefen automatisch und die persönlichen Ergebnisse der Spenderinnen und Spender wurden ebenfalls automatisch an unseren Datenspende-Server geschickt. Pro Nutzer und Zeitpunkt wurden also zweimal 16 Suchbegriffe abgefragt und die jeweils erste Seite der Suchergebnisse gespeichert.

Die Suchbegriffe beschränkten sich auf die sieben großen Parteien und ihre jeweiligen Parteispitzen (siehe Tabelle 1Tabelle 2). Wie in Abbildung 1 zu sehen ist, konnten die Nutzer nach dem Herunterladen des Plug-ins frei entscheiden, ob sie über zukünftige Spenden informiert werden wollten oder ob diese weitestgehend im Hintergrund ablaufen sollten.

[5] https://github.com/algorithmwatch/datenspende, veröffentlicht am 7.6.2017.

[6] Sollte der Browser zu einem oder mehreren aufeinanderfolgenden Suchzeitpunkten ausgeschaltet gewesen sein, wurde beim nächsten Einschalten direkt eine Runde von Suchanfragen gestartet – daher gibt es auch Suchergebnislisten mit abweichenden Zeitstempeln. Zudem war es manuell möglich, im eigenen Browser einen Suchzyklus auszulösen.

Tabelle 1: Suchbegriffe des Datenspende-Plug-Ins zu festgelegten Zeitpunkten.

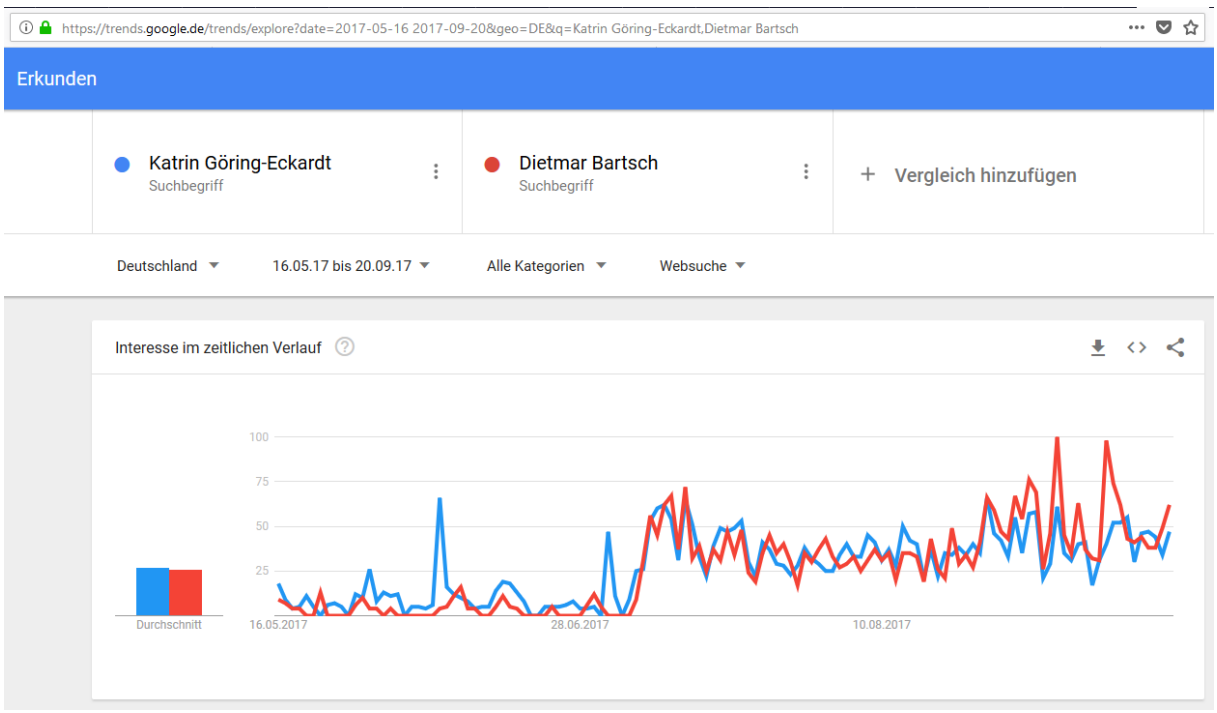
Suchbegriffe: Parteien
AFD
Bündnis90/Die Grünen
CDU
CSU
Die Linke
FDP
SPD

Suchbegriffe: Personen
Alice Weidel
Dietmar Bartsch
Alexander Gauland
Katrin Göring-Eckardt
Christian Lindner
Angela Merkel
Cem Özdemir
Martin Schulz
Sahra Wagenknecht

Die Informationen über das Projekt und der damit verbundene Aufruf zur Datenspende wurden sowohl über Kommunikationskanäle der Projektpartner als auch über unseren Medienpartner Spiegel Online verbreitet (Horchert, 2017). Daraufhin erfolgten 4.384 Installationen des Plug-Ins. Die hieraus resultierten Suchergebnisse sind der Bevölkerung zu Analysezwecken frei zugänglich[7].

An dieser Stelle muss darauf hingewiesen werden, dass alle Ergebnisse, von denen in diesem Abschlussbericht zu lesen ist, nicht notwendigerweise repräsentativ sind, da sich die Datenspenderinnen und -spender freiwillig und selbstselektiert zu einer Teilnahme entschlossen haben. Bei den wichtigen Ergebnissen insbesondere zum Personalisierungsgrad gehen wir allerdings davon aus, dass diese sich nicht wesentlich verändern, wenn die Nutzerbasis repräsentativ ist.

Es muss weiterhin bemerkt werden, dass die automatische Suche nach circa einem Dutzend Suchbegriffen auch Einfluss auf den Suchmaschinenalgorithmus selbst haben kann. Auf Google Trend kann man für die Suchbegriffe "Dietmar Bartsch", "Katrin Göring-Eckardt" und "Bündnis90/Die Grünen" gut erkennen, dass über die Laufzeit der Datensammlung das Volumen der Suchanfragen dadurch erhöht wurde (siehe Abbildung 2).



[7] https://datenspende.algorithmwatch.org/data.html

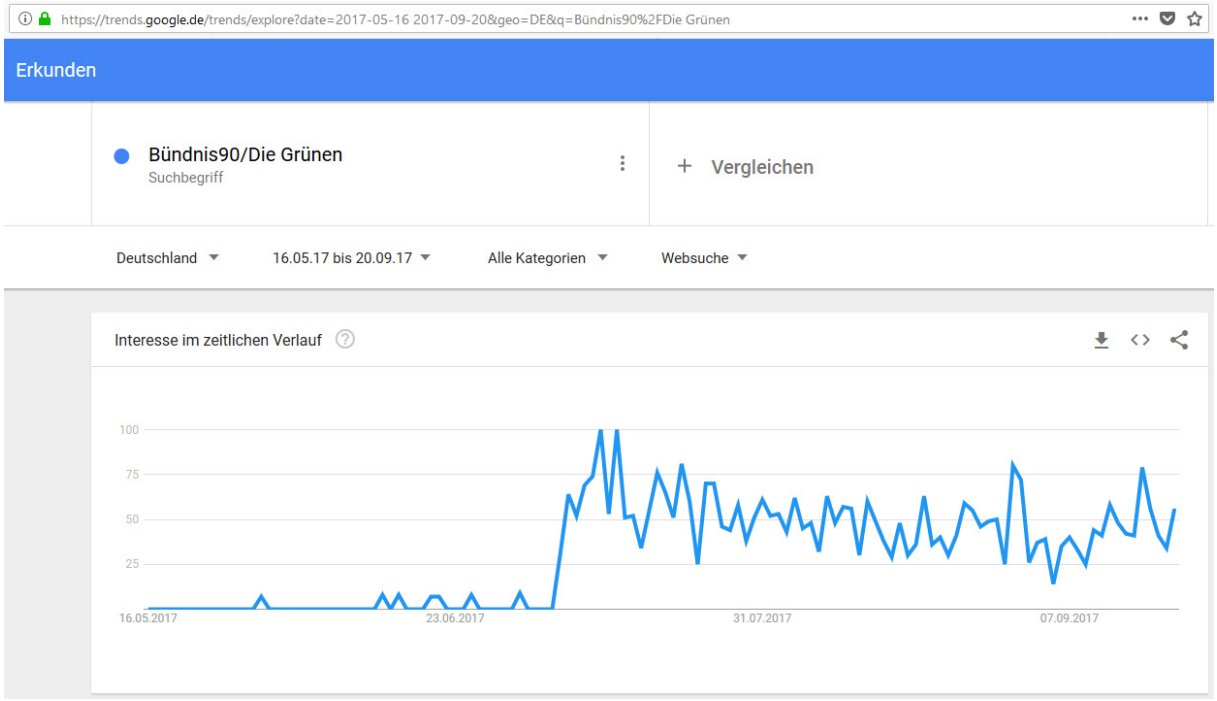


Abbildung 2: Zeitlicher Verlauf der Suchbegriffe „Katrin Göring-Eckardt“ und „Dietmar Bartsch“ (oben) und des Suchbegriffes „Bündnis90/Die Grünen“ (unten). Die Diagramme von Google-Trends zeigen deutlich die Erhöhung des Suchaufkommens der Begriffe aufgrund des Plug-Ins, das am 6.7.2017 freigeschaltet wurde.

Da die Suchanfragen aber automatisch gestellt wurden und auf keinen der angebotenen Links aktiv geklickt wurde, vermuten wir, dass die Auswirkungen eher gering sind. Dies lässt sich allerdings ohne genaue Kenntnis des dahinterliegenden Algorithmus nicht nachweisen und muss daher unbewertet bleiben.

2.2 Struktur der Daten und wichtige Begriffe

Die zur Verfügung stehenden Daten wurden zur Untersuchung wie folgt strukturiert. Zum einen wurde unterschieden, ob es sich um Suchergebnisse direkt auf Google (www.google.de) oder auf der Nachrichtenseite des Suchmaschinenbetreibers (news.google.com) handelt. Der gesamte zur Verfügung stehende Datenbestand wurde in diese beiden Bereiche aufgeteilt. Während auf dem Nachrichtenportal von Google in der Regel 20 dem Suchbegriff entsprechende Treffer angezeigt werden, werden dem Nutzer bei der Suche auf der regulären Suchmaschinen von Google neben in der Regel zehn Treffern auch bis zu drei sogenannte Schlagzeilen angezeigt (siehe Abbildung 3).

Während Google-Suchen hauptsächlich auf persönliche Webseiten, Social-Media-Accounts und Aggregator-Themenseiten von Parteien und Personen verweisen (vgl. Abschnitt 3.4), zeigt die Google-News-Suche nur Nachrichten von vorher registrierten Partnern an[8]. Die Lebensdauer von Nachrichten ist dabei eher begrenzt, d. h. die meisten Nachrichten werden den Nutzern nur über wenige Suchzeitpunkte angeboten, während die persönlichen Webseiten und Social-Media-Accounts der Parteien und Politikerinnen und Politiker fast immer angezeigt werden. Insofern ist eine getrennte Betrachtung der Ergebnisse der News-Suche und der Google-Suche sinnvoll.

Bei den Ergebnissen der normalen Google-Suche haben wir zusätzlich zwischen den (nicht immer ausgelieferten) Schlagzeilen und den „organischen“ Suchergebnissen unterschieden, d. h., den 8-10 Ergebnissen im linken unteren Teilfeld der Ergebnisseite (Abbildung 3 zeigt zwei der organischen Suchergebnisse zum Suchbegriff „Angela Merkel“). Manchmal enthalten Suchergebnisseiten von Google im rechten Teilfeld ebenfalls Informationen, z.B. Anzeigen oder Infoboxen über Personen und Parteien. Diese

[8] Siehe die Hilfeseiten zu Google News, z.B. https://support.google.com/news/publisher-center/answer/6016113?hl=de&ref_topic=9010378.

wurden nicht an unsere Server übertragen, sondern nur eventuelle Schlagzeilen und die eigentlichen Suchergebnisse.

Neben dem Typ der Suche („Google Suche“, „Google News Suche“) wurden gespeichert:

- Eine ungefähre Standortbestimmung aufgrund der übermittelten IP-Adresse des Nutzers.
- Der Login-Status der Nutzer auf ihrem Google-Account. Ein Nutzer kann "eingeloggt" oder "nicht eingeloggt" sein.
- Die im Browser eingestellte Sprache (nicht die Suchsprache, die man auf dem Google-Account eingibt).
- Der Suchbegriff und der Zeitstempel der Suche.
- Eine vom Plug-In generierte ID, die ansonsten keine Hinweise auf den Nutzer oder die Nutzerin gibt, aber für alle gesammelten Datenspenden gleichbleibt, solange das Plug-in nicht neu installiert wird.
- Soweit verfügbar, wurde ein beschreibender Text der Links mitgespeichert (meistens nur für organische Suchergebnisse verfügbar, aber nicht immer).
- Wenn es sich um eine Schlagzeile handelt, wurde die entsprechende Zeitangabe (z.B.: "Vor 54 Minuten", "Vor 3 Stunden") mitgespeichert.
- Die URL des Suchergebnisses.
- Bei Schlagzeilen und Ergebnissen der Google News Suche wird meistens noch das Medium (die Nachrichtenquelle) und ein Titel der Nachricht mitangegeben und von uns gespeichert (z.B. "Dresdner Neueste Nachrichten" mit "Das sagen unsere Leser zum Auftritt von Cem Özdemir").

Im vorliegenden Bericht werden zudem folgende Begriffe verwendet:

Untersuchungszeitraum: Als Untersuchungszeitraum bezeichnen wir den Zeitraum vom 21.08.2017 bis zum 24.09.2017. Hierbei sind lediglich die Wochentage sowie als einziges Wochenende das Wahlwochenende berücksichtigt (näheres hierzu siehe Abschnitt 3.1). Der Untersuchungszeitraum beinhaltet somit 27 Tage.

Suchzeitpunkt: Unter einem Suchzeitpunkt verstehen wir einen im Untersuchungszeitraum liegenden Tag und die dazugehörige Uhrzeit, die 12:00 Uhr, 16:00 Uhr oder 20:00 Uhr betragen kann. Wir beschränken uns auf diese Zeitpunkte, da zu den anderen Suchzeitpunkten deutlich weniger Nutzer suchen. Die Gesamtheit der Suchzeitpunkte beträgt somit 81 (jeweils drei unterschiedliche Zeitpunkte an insgesamt 27 Tagen).

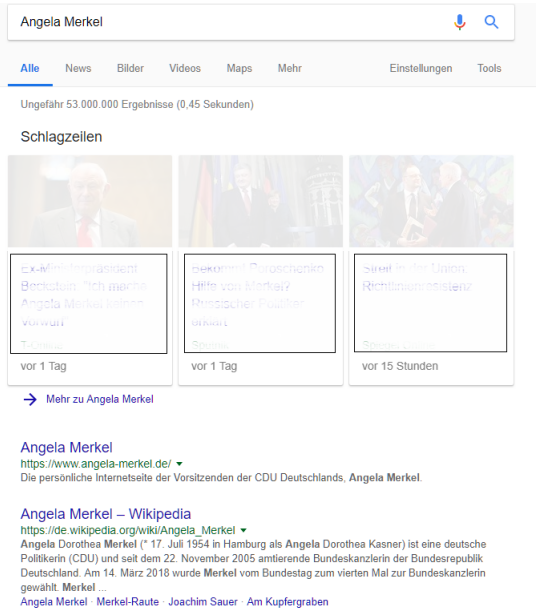


Abbildung 3: Google Suche nach Angela Merkel mit drei ausgelieferten Schlagzeilen.

(Such-) Ergebnisliste: Unter einer Ergebnisliste verstehen wir die Menge der URLs, die zu einem gegebenen Suchbegriff und einem definierten Suchzeitpunkt an einen Nutzer ausgeliefert wurden.

Schlagzeilen und organische Suchergebnisse: Als Schlagzeilen werden die bis zu drei Nachrichtenmeldungen bezeichnet, die Google manchmal (siehe Abschnitt 6.1) an oberster Position einer regulären Google Suchanfrage ausliefert. Zu diesen wird neben der reinen textuellen Information noch eine entsprechende Abbildung angezeigt (siehe Abbildung 3). Die restlichen Suchergebnisse werden im Folgenden als organische Suchergebnisse bezeichnet.

Top-Level-Domain: Jede URL verweist auf eine Hauptdomain (Top-Level-Domain) und unter Umständen dort liegende Verzeichnisse. Die Hauptdomain entspricht dem Teil zwischen der Protokollangabe ("http:// ", https://) der URL bis zum ersten Slash ("/"). Die Top Level-Domain von http://www.faz.net/aktuell/wirtschaft/gruenen-chef-cem-oezdemir-will-gelaendewagen-bestrafen-wer-suv-faehrt-soll-die-kosten-fuer-die-umwelt-tragen-15201893.html ist somit www.faz.net.

Wie bei allen Datensammlungen kommt es auch zu Einträgen, die fehlerhaft oder anders als die Norm sind. Auch das Plug-In ist nicht von Anfang an rund gelaufen und produzierte teilweise fehlerhafte Daten. Daher beschreiben wir im nächsten Abschnitt die notwendige Datenaufbereitung.

2.3 Datenaufbereitung

Das erste Plug-In der Firefox-Version wies allen Nutzern dieselbe ID zu. Da wir auch Analysen der Veränderung von Suchergebnislisten über die Zeit vornehmen wollten, haben wir uns für die hier vorliegenden Analysen dafür entschieden, diese Daten grundsätzlich nicht zu berücksichtigen, um eine einheitliche Datenbasis zu haben. Damit fallen 34 % aller gespendeten URLs bei der Google Suche weg und 41 % bei den Google-News-Suchen.

Eine erste Analyse der zur Verfügung stehenden Daten (siehe Abschnitt 3) zeigte weitere Unregelmäßigkeiten. Auffällig war zum Beispiel, dass der Datenbestand Suchergebnislisten enthielt, die in der Länge nicht den erwarteten Standards (zehn Einträge bei der reinen Google-Suche und 20 Einträge bei der Suche auf dem Nachrichtenportal) entsprachen. Die Datenbasis enthält beispielsweise einige Datensätze mit 200 Einträgen in den Ergebnislisten. Dies ist auf die Möglichkeit zurückzuführen, im Google-Account die Anzahl der auf der ersten Seite angezeigten Suchergebnisse einzustellen. Wir haben diese Listen auf die üblichen 10 plus eventuelle angezeigte Schlagzeilen gekürzt. Andere Fehler beruhen auf einer fehlerhaften Programmierung des ersten Firefox-Plug-Ins und erzeugten damit Suchergebnislisten, in denen überall dieselbe URL stand. Diese Listen wurden nicht mit in die Analyse einbezogen.

Gleiches gilt für URLs, die lediglich einen Verweis auf die entsprechende URL bei Google (google.de/url) enthielten oder einen URL-Eintrag aufwiesen, der lediglich auf „google“ verweist, also keinen ganzen Link enthielt – diese Einträge weisen auf ein privates Suchergebnis hin. Weiterhin war auffällig, dass eine Reihe von Suchergebnislisten größere Anzahlen von URLs enthielten, die auf anderssprachige Webseiten verwiesen.

Um ausländische Ergebnisse auszusortieren, haben wir manuell Top-Level-Domains aller Webseiten nach Sprachen sortiert. Damit haben wir für jede Suchergebnisliste den Anteil deutschsprachiger Webseiten ermittelt und behielten die, deren Anteil über 50% lag. Unsere Datenbasis wurde somit auf deutschsprachige Ergebnislisten eingeschränkt. Die Datensätze im Untersuchungszeitraum für die Google-Suche wurden durch diese Reinigungsschritte um 19,6 %[9] und die Google News-Daten um 16,7 %[10] reduziert. Eine kurze Aufstellung über die bereinigten Datensätze befindet sich inTabelle 2.

Tabelle 2: Tabellarische Übersicht über die Datenspende-Daten nach der Datenaufbereitung.

	Bereinigter Datenbestand Google Suche	Bereinigter Datenbestand Google News
Anzahl der Datensätze	3.564.583	5.597.480
Anzahl Nutzer	1.705	1.550
Anzahl Ergebnislisten	302.813	279.874

[9] Von 4.416.585 auf 3.560.542
[10] Von 6.712.733 auf 5.597.480

3 Überblick über die Daten

In diesem Abschnitt geben wir einen ersten Überblick über die Daten.

3.1 Zeitliche Verteilung der Suchanfragen

Bei der Betrachtung der Tagesverteilung unserer Datenspenden zeigte sich ein deutlicher Rückgang der gespendeten Suchergebnislisten während der Wochenenden. In Abbildung 4 ist die Anzahl eingegangener URLs für die Google-Suche pro Tag und Suchbegriff dargestellt und es zeigt sich ein wellenartiges Muster, wobei die Tiefpunkte an den Wochenenden liegen. Es haben also an den Wochenenden deutlich weniger Nutzerinnen und Nutzer ihre Browser geöffnet gehabt und uns Daten gespendet, was wahrscheinlich auf eine geringere Computernutzung am Wochenende zurückgeht.

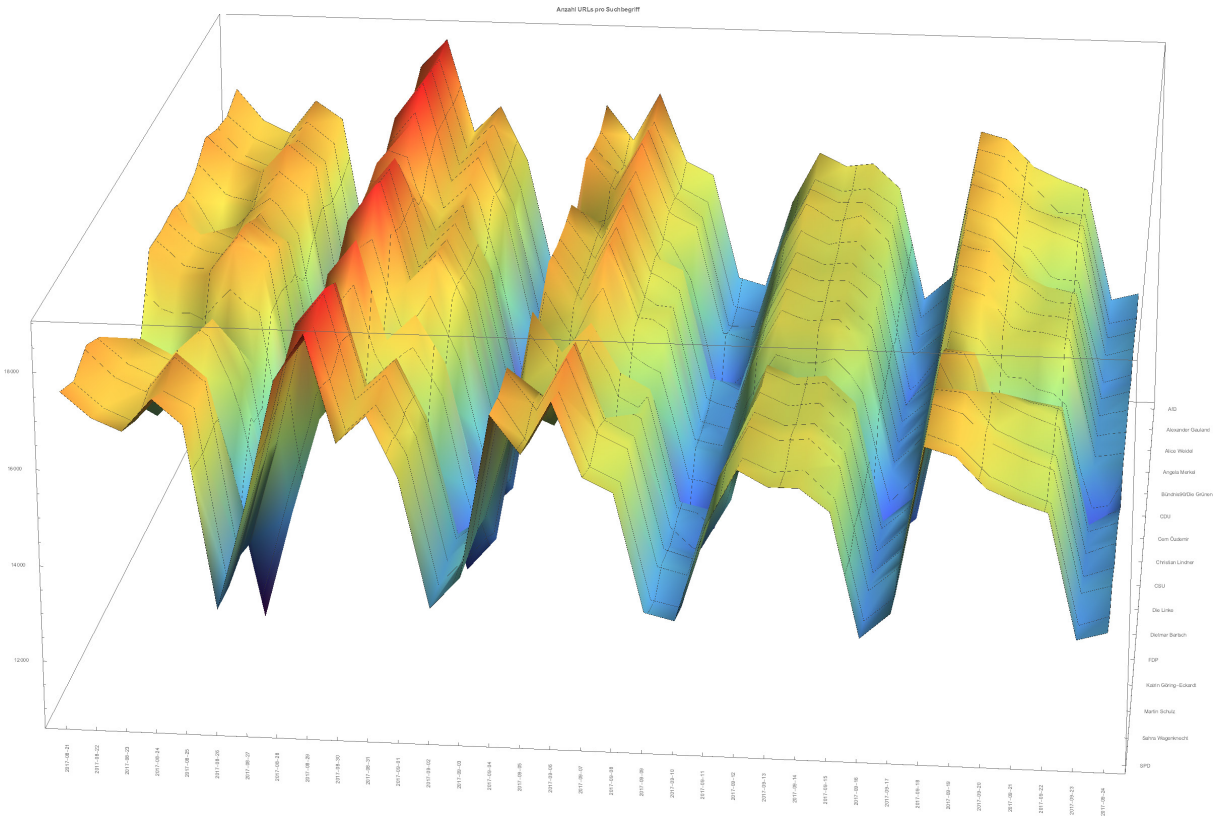


Abbildung 4: Anzahl eingereichter URLs bei der Google Suche während der letzten 5 Wochen vor der Bundestagswahl 2017. Auf der X- Achse sind die Suchzeitpunkte, auf der Y-Achse die Anzahl der aktiven Datenspender und auf der Z-Achse die Suchbegriffe aufgetragen.

Die tägliche Anzahl an Datenspendern konnte durch das Einschränken auf die Wochentage über den Untersuchungszeitraum sowohl für die Google Suche (siehe Abbildung 5) als auch bei Google News (siehe Abbildung 6) recht stabil gehalten werden. So waren zu den jeweiligen Tagen stets zwischen 450 und 550 Nutzer online und haben uns Ihre Suchergebnisse gespendet. Auch blieb das Verhältnis von eingeloggten zu nicht eingeloggten Nutzern über den Untersuchungszeitraum ähnlich. Das Wochenende vom 23.9./24.09.2018 wurde hierbei lediglich durch seine Nähe zur Wahl hinzugenommen und fällt mit seinen Nutzerzahlen von ca. 350 Nutzern etwas aus dem Rahmen. Es sind aber statistisch gesehen immer noch genügend große Fallzahlen, um diese Tage mit zu betrachten.

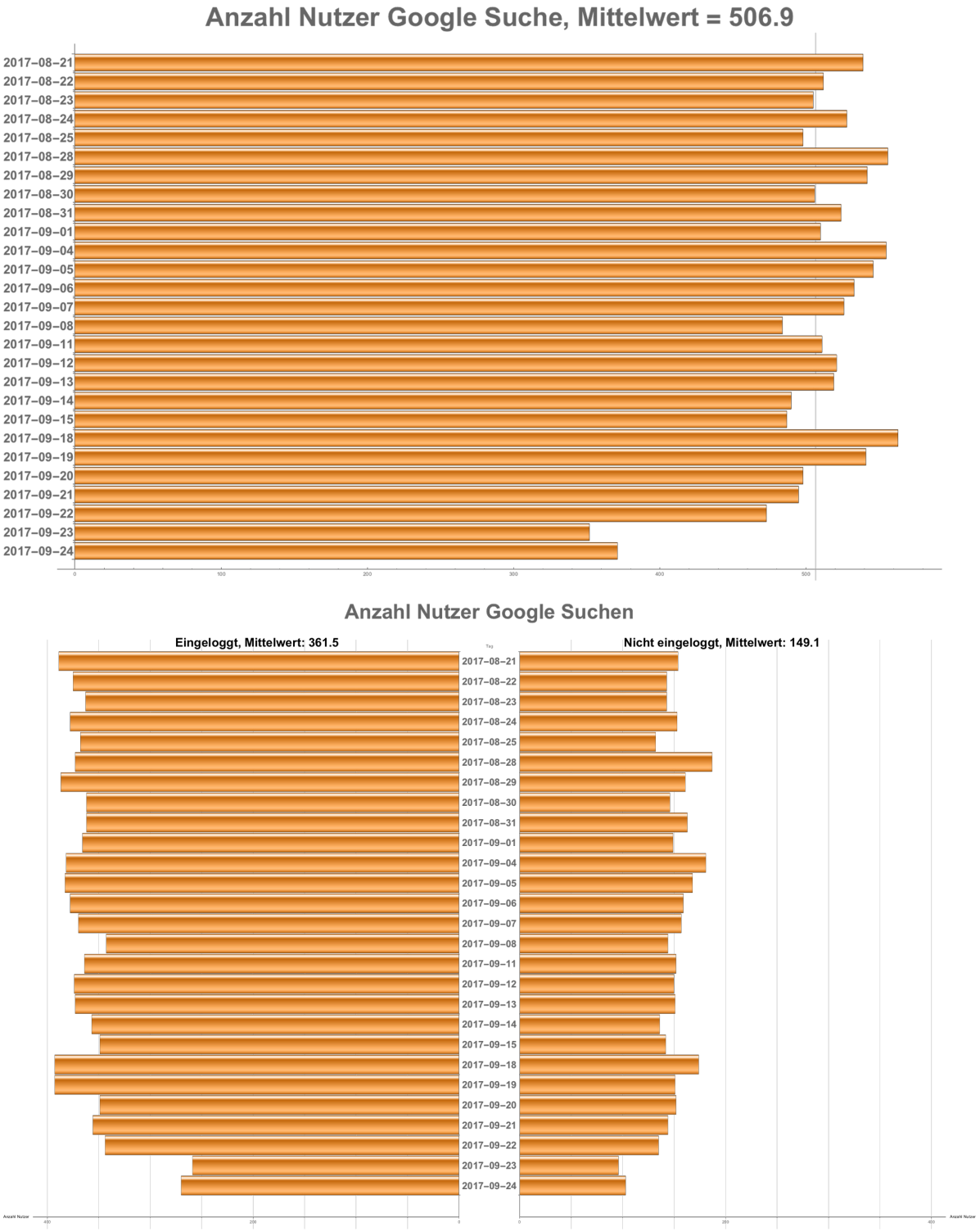


Abbildung 5: Anzahl der Nutzer an den jeweiligen Tagen des Untersuchungszeitraums bei der Google-Suche, nach der Bereinigung der Daten.

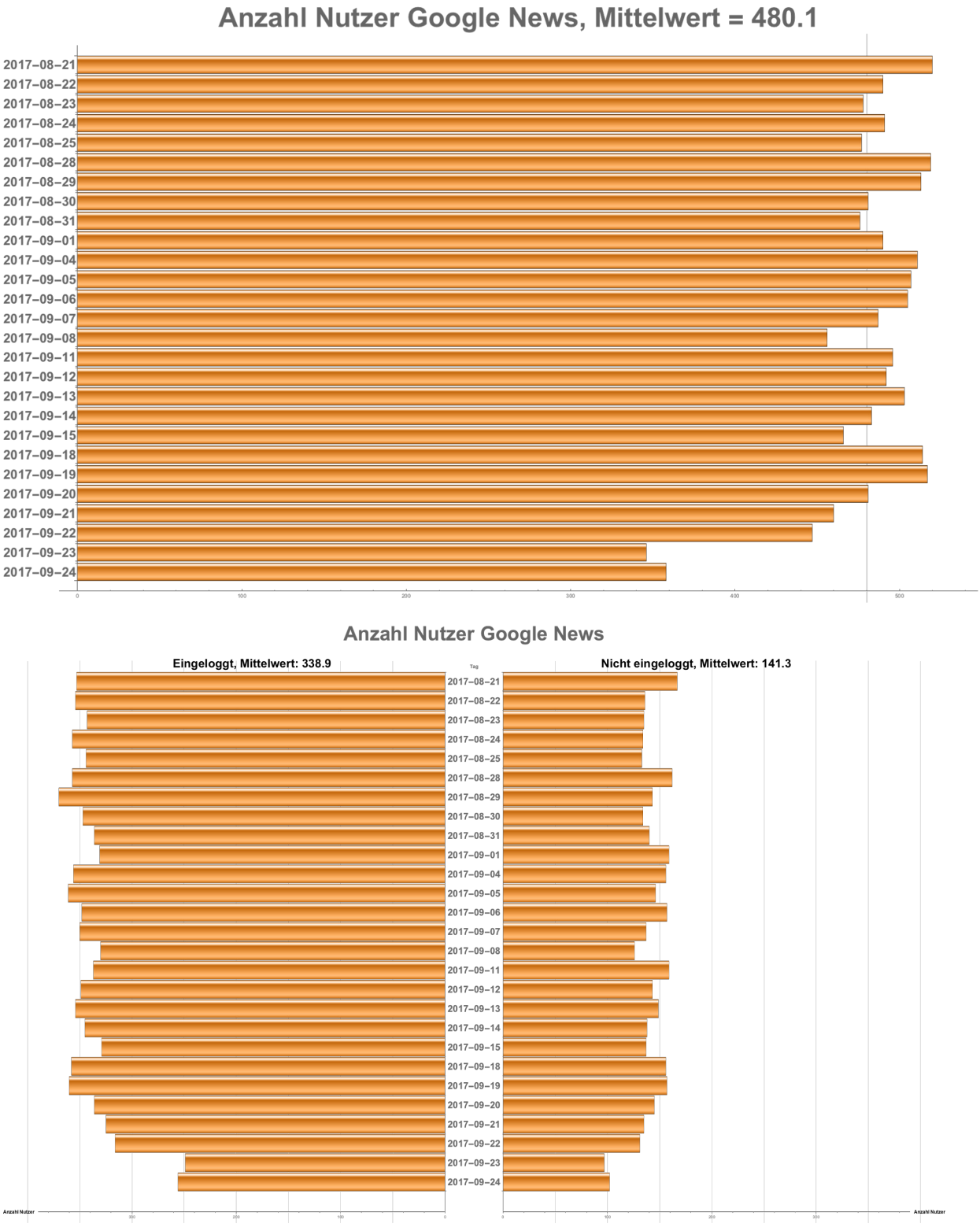


Abbildung 6: Anzahl der Nutzer an den jeweiligen Tagen des Untersuchungszeitraums bei Google News nach der Bereinigung der Daten.

3.2 Geographische Verteilung der Datenspenderinnen und -spender

Die Verteilung der Nutzer über die Deutschlandkarte in Abbildung 7 zeigt zum einen, dass wir es geschafft haben, deutschlandweit Datenspender zu akquirieren und anderseits unsere Datenaufbereitung (siehe Abschnitt 2.3) auf keine regionalen Effekte schließen lässt: Orte, die nach der Datenbereinigung nicht mehr enthalten sind, werden als rote Punkte dargestellt. Diese verteilen sich über ganz Deutschland; insgesamt ist der Westen Deutschlands stärker vertreten.

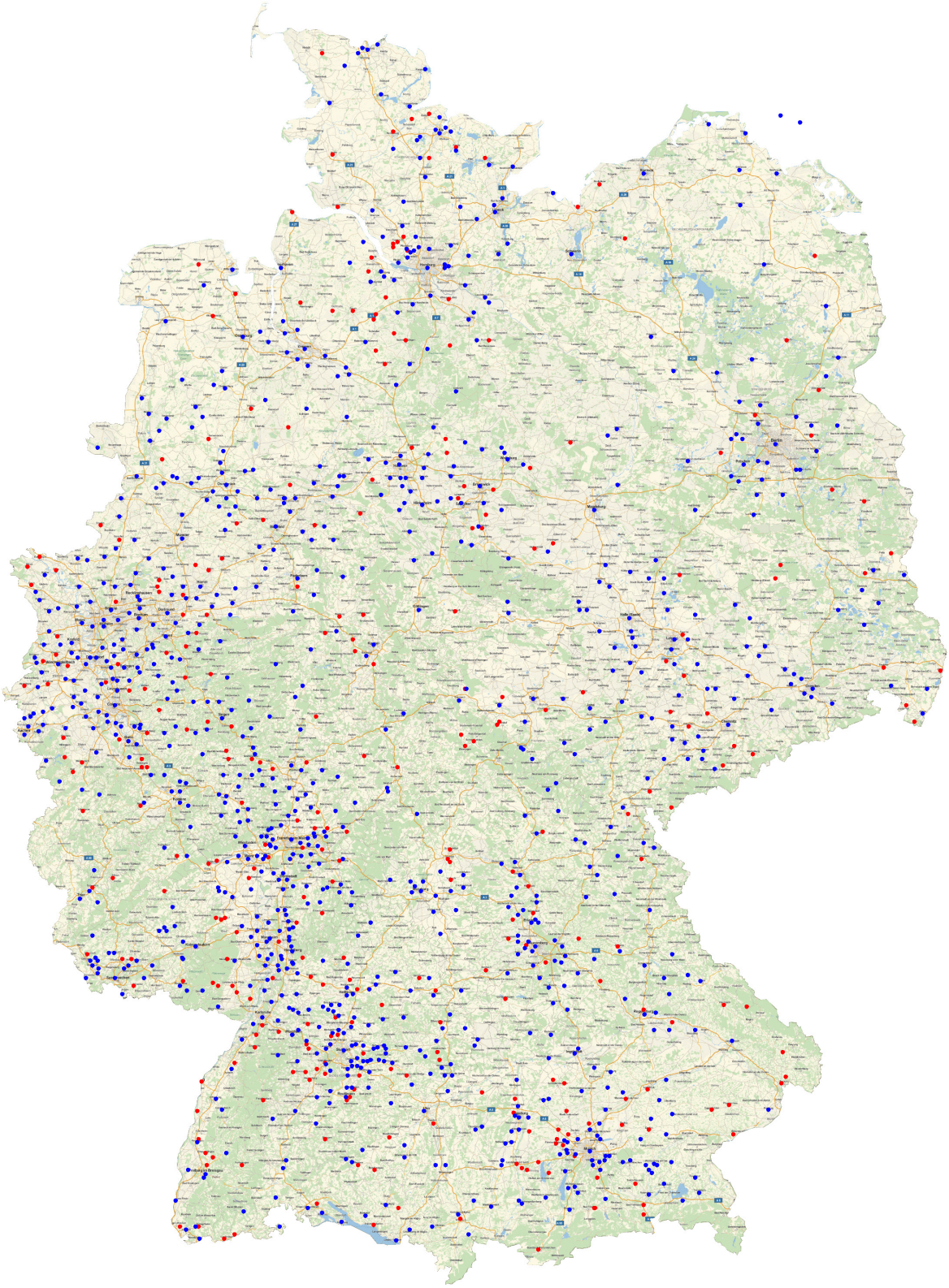


Abbildung 7: Verteilung der Datenspender über Deutschland, hierbei wurden die jeweiligen Positionen auf Grundlage der IP-Adresse ermittelt und sind deshalb nur eine Näherung an die Position der einzelnen Nutzer. Die roten Punkte sind hierbei Koordinaten, die nach dem Reinigungsschritt nicht mehr in der Datenbasis enthalten sind.

3.3 Verteilung von Internet- und Medienangeboten in den Suchergebnissen

Für den verbliebenen Datensatz haben wir zuerst berechnet, welche Top-Level-Domains es insgesamt am häufigsten pro Suchbegriff auf die erste Suchergebnisseite geschafft haben. Abbildung 8 zeigt die Ergebnisse für die Parteien und Abbildung 9 für die Politikerinnen und Politiker. Bei allen Parteien außer für „Die Linke“ steht die jeweilige parteieigene Top Level-Domain an erster Stelle, bei „Die Linke“ erst an vierter Stelle, dafür steht die Webseite der Linksfraktion im Bundestag an zweiter Stelle. Der jeweilige Facebook-Account ist ebenfalls immer der zweit- bis vierthäufigste Link, Twitter schafft es bei vier Parteien in die Top 10; dazu gesellt sich bei allen Parteien außer der AfD auch der deutsche Wikipedia-Eintrag der Partei. Bei der AfD gibt es zwar auch einen deutschen Wikipedia-Eintrag, dieser taucht nur nicht so häufig auf wie die Themen-Aggregationsseiten von Focus, Zeit, Merkur, Spiegel, FAZ und Tagesspiegel. Ebenfalls oft zu sehen für alle Parteien außer den Grünen ist die Top-Level-Domain www.bundestagswahl-bw.de, die einen Überblick über alle Parteien bildet. Die bei den Grünen und der SPD auftauchende Seite „bundestagswahl-2017.com“ ist übrigens eine private Informationssammlung über die Parteien und deren Parteiprogramme. Es fällt auf, dass „Bündnis90/Die Grünen“ und „Die Linke“ in ihren Top-10-Ergebnissen mehrheitlich eigene Webseiten haben oder solche, die von ihnen mit einigem Aufwand verändert werden können, falls es zu falschen Berichterstattungen kommt. Dazu zählen wir auch Facebook und Twitter sowie grundsätzlich auch die Wikipedia.

AfD

www.afd.de	18 431
www.welt.de	16 347
www.bundestagswahl-bw.de	15 285
de-de.facebook.com	14 102
www.focus.de	12 939
www.zeit.de	11 730
www.merkur.de	9955
www.spiegel.de	8398
www.faz.net	7384
www.tagesspiegel.de	6845

Gesamtanzahl an URLS: 164980

Bündnis90/Die Grünen

www.gruene.de	41 830
de-de.facebook.com	20 000
de.wikipedia.org	18 560
www.gruene-bundestag.de	17 420
bundestagswahl-2017.com	13 888
gruene.berlin	5882
www.bpb.de	4244
www.youtube.com	3561
gruene-nrw.de	3097
twitter.com	2824

Gesamtanzahl an URLS: 169114

CDU

www.cdu.de	18 922
de.wikipedia.org	18 332
de-de.facebook.com	17 806
www.bundestagswahl-bw.de	14 674
www.zeit.de	11 381
twitter.com	10 685
www.welt.de	8568
www.focus.de	5223
www.spiegel.de	5210
www.faz.net	4298

Gesamtanzahl an URLS: 164374

CSU

www.csu.de	21 427
de.wikipedia.org	17 706
twitter.com	15 971
de-de.facebook.com	14 635
www.zeit.de	13 839
www.spiegel.de	12 360
www.welt.de	7766
www.bundestagswahl-bw.de	5874
www.focus.de	5022
www.tagesspiegel.de	4412

Gesamtanzahl an URLS: 152764

Die Linke

de.wikipedia.org	19 031
www.linksfraktion.de	18 661
de-de.facebook.com	13 837
die-linke.de	12 879
twitter.com	12 526
www.die-linke.de	8274
www.bundestagswahl-bw.de	8154
bundestagswahl-2017.com	6335
www.faz.net	5513
www.youtube.com	4488

Gesamtanzahl an URLS: 152339

FDP

www.fdp.de	27 278
de.wikipedia.org	18 505
de-de.facebook.com	15 539
twitter.com	13 803
www.welt.de	11 759
www.bundestagswahl-bw.de	7825
www.liberale.de	4650
www.focus.de	2826
www.faz.net	2728
www.fdp-berlin.de	2081

Gesamtanzahl an URLS: 152862

SPD

www.spd.de	29 201
de.wikipedia.org	19 024
www.bundestagswahl-bw.de	17 461
de-de.facebook.com	15 830
www.welt.de	12 683
bundestagswahl-2017.com	11 476
www.zeit.de	8676
www.faz.net	6905
www.focus.de	6088
www.spiegel.de	4622

Gesamtanzahl an URLS: 178017

Abbildung 8: Häufigste Top-Level-Domains bei der organischen Google-Suche nach den Parteien.

Für die Politiker ist ebenfalls immer die persönliche Webseite unter den Top 10, bei Alexander Gauland und Alice Weidel sind dies Unterseiten der Top-Level-Domain www.afd.de und daher nicht eigens sichtbar. Bei allen ist auch der entsprechende deutsche Wikipedia-Eintrag zu finden, neben einer wechselnden Anzahl von Social-Media-Accounts, die nur bei Alexander Gauland völlig fehlen. Tatsächlich konnten wir auch keinen persönlichen Social-Media-Account von Alexander Gauland finden, seine Webseite verweist auf die jeweiligen Accounts der AfD.

Die meisten großen deutschen Online-Nachrichtenmagazine haben auch Themenseiten, auf denen Nachrichten z. B. über eine Person oder Institution gesammelt werden und oft auch ein einleitender, allgemeiner Text über das Thema zu finden ist. Solche Themenseiten zu den entsprechenden Personen finden sich bei den Politikern und Politikerinnen unter den Top 10 der Top Level-Domains, aber durchaus auch Nachrichten von diesen und anderen Quellen.

Alexander Gauland

www.welt.de	26 366
www.spiegel.de	26 330
www.faz.net	23 340
de.wikipedia.org	19 747
www.zeit.de	17 907
afd-fraktion-brandenburg.de	7503
web.de	7492
www.t-online.de	6421
www.afd.de	5719
www.bild.de	4534

Gesamtanzahl an URLS: 179070

Alice Weidel

www.faz.net	21 231
www.afd.de	19 538
de.wikipedia.org	19 472
twitter.com	17 343
uebermedien.de	14 599
www.welt.de	10 474
de-de.facebook.com	10 013
www.focus.de	9806
www.facebook.com	9698
www.bild.de	6683

Gesamtanzahl an URLS: 176493

Angela Merkel

www.spiegel.de	22 280
www.bundeskanzlerin.de	20 320
www.focus.de	18 851
de.wikipedia.org	17 611
www.angela-merkel.de	17 544
de-de.facebook.com	17 132
www.zeit.de	10 735
www.welt.de	7365
www.instagram.com	5974
www.t-online.de	4411

Gesamtanzahl an URLS: 159040

Cem Özdemir

web.de	21 626
twitter.com	20 031
www.oezdemir.de	19 671
de.wikipedia.org	19 662
de-de.facebook.com	17 866
www.bundestag.de	16 442
www.faz.net	15 661
www.gruene.de	8433
www.spiegel.de	7250
www.zeit.de	4460

Gesamtanzahl an URLS: 178748

Christian Lindner

www.christian-lindner.de	23 785
twitter.com	19 341
de.wikipedia.org	19 258
www.welt.de	18 970
de-de.facebook.com	17 533
www.spiegel.de	17 316
www.faz.net	7129
www.zeit.de	4644
www.focus.de	4530
www.huffingtonpost.de	4014

Gesamtanzahl an URLS: 174977

Dietmar Bartsch

www.dietmar-bartsch.de	19 992
de.wikipedia.org	19 991
twitter.com	19 983
www.bundestag.de	19 923
de-de.facebook.com	18 154
www.spiegel.de	17 436
web.de	17 081
www.stern.de	14 456
www.faz.net	12 456
www.linksfraktion.de	10 141

Gesamtanzahl an URLS: 184959

Katrin Göring-Eckardt

www.goering-eckardt.de	20 933
www.spiegel.de	19 956
de.wikipedia.org	19 956
www.focus.de	19 091
www.bundestag.de	19 063
twitter.com	17 810
de-de.facebook.com	17 263
web.de	14 168
www.zeit.de	7879
www.welt.de	7288

Gesamtanzahl an URLS: 183340

Martin Schulz

www.spiegel.de	20 042
martinschulz.de	18 675
twitter.com	18 510
de.wikipedia.org	18 434
de-de.facebook.com	16 811
www.faz.net	12 776
www.zeit.de	11 268
www.welt.de	8765
www.focus.de	7332
www.bild.de	6101

Gesamtanzahl an URLS: 168050

Sahra Wagenknecht

www.spiegel.de	19 591
www.sahra-wagenknecht.de	19 433
de.wikipedia.org	19 329
de-de.facebook.com	17 474
www.welt.de	15 671
www.faz.net	13 486
www.tagesspiegel.de	13 351
www.bundestag.de	10 777
twitter.com	10 577
www.linksfraktion.de	8703

Gesamtanzahl an URLS: 176105

Abbildung 9: Häufigste Top Level-Domains bei der organischen Google Suche nach den Personen.

Untersucht man die Top Level-Domains bei den Schlagzeilen ergibt sich ein weniger einheitliches Bild (siehe Abbildung 10): Während die AfD, CDU, CSU, SPD und FDP am meisten Schlagzeilen aus den reichweitenstarken Medienhäusern bekommen, sind die Quellen unter den Top 10 bei den Grünen und Linken teilweise weniger bekannt. Auch bei den gesuchten Personen sind neben den reichweitenstarken Quellen auch andere zu finden, wie der Generalanzeiger Bonn (für Cem Özdemir) oder www.epochtimes.de (für Alexander Gauland und Katrin Göring-Eckhardt) (siehe Abbildung 11). Der Vollständigkeit halber zeigen Abbildung 12 und Abbildung 13 auch die Top 10 der Top Level-Domains bei den jeweiligen Suchen auf Google News. Eine medienwissenschaftliche Einordnung der jeweiligen

Quellen kann hier nicht geliefert werden – die Daten stehen aber für weitere Analysen öffentlich zur Verfügung[11].

Insgesamt werden die erste Ergebnisseite bei der Suche nach Nachrichten mit Google News und die von Google im Rahmen der organischen Suche ausgerollten Schlagzeilen von etablierten Medienhäusern, insbesondere aus dem Printbereich, dominiert. Ausnahmen stellen vor allem die Online-Zeitung Huffington Post sowie das Freemail Portal t-online.de dar, die häufig als Nachrichtenquellen auftauchen. Bei der Suche nach den Begriffen „Christian Lindner“ und „FDP“ liefert Google mit liberale.de eine Website als Nachrichtentreffer aus, die von der Bundesgeschäftsstelle der Partei verantwortet wird.

AfD

www.welt.de	11 014
www.focus.de	6091
www.zeit.de	5954
www.spiegel.de	5839
www.faz.net	4903
www.t-online.de	3847
www.bild.de	3651
www.merkur.de	2542
www.sueddeutsche.de	2070
www.tagesspiegel.de	1498

Gesamtanzahl an URLS: 54122

Bündnis90/Die Grünen

www.mz-web.de	3036
www.tagesspiegel.de	2898
t3n.de	2650
www.neustadt-ticker.de	1873
www.chiemgau24.de	1738
www.bild.de	1719
www.heise.de	1695
www.deutschlandfunk.de	1526
deutsch.rt.com	1502
www.handelsblatt.com	1374

Gesamtanzahl an URLS: 47700

CDU

www.welt.de	11 543
www.focus.de	7458
www.faz.net	4831
www.spiegel.de	4608
www.zeit.de	2476
www.epochtimes.de	2376
www.sueddeutsche.de	1728
www.t-online.de	1715
www.bild.de	1676
web.de	1622

Gesamtanzahl an URLS: 52819

CSU

www.sueddeutsche.de	12 200
www.welt.de	6343
www.focus.de	5205
www.zeit.de	4087
www.spiegel.de	3631
www.n-tv.de	2312
www.merkur.de	2014
www.bild.de	1671
www.epochtimes.de	1526
www.t-online.de	1219

Gesamtanzahl an URLS: 51113

Die Linke

www.welt.de	7374
www.faz.net	4775
www.epochtimes.de	4426
www.zeit.de	3308
www.bento.de	3042
www.tagesspiegel.de	2483
www.focus.de	2305
www.tichyseinblick.de	2054
www.ndr.de	1394
www.n-tv.de	1378

Gesamtanzahl an URLS: 53523

FDP

www.welt.de	5666
www.focus.de	4126
www.spiegel.de	3752
www.faz.net	3704
www.liberale.de	3647
www.sueddeutsche.de	2635
www.n-tv.de	2505
www.stern.de	2488
www.tagesspiegel.de	2074
www.t-online.de	1977

Gesamtanzahl an URLS: 52535

SPD

www.welt.de	13 620
www.faz.net	7250
www.focus.de	6040
www.zeit.de	5425
www.spiegel.de	5128
www.t-online.de	4417
www.sueddeutsche.de	3247
www.n-tv.de	2797
www.bild.de	2245
www.stern.de	1033

Gesamtanzahl an URLS: 59058

Abbildung 10: Häufigste Top-Level-Domains bei den Schlagzeilen zur Google-Suche nach den Parteien

[11] https://datenspende.algorithmwatch.org/data.html

Alexander Gauland

www.welt.de	12 164
www.zeit.de	6526
www.spiegel.de	4971
www.faz.net	4585
www.focus.de	2717
www.epochtimes.de	2638
www.t-online.de	2436
www.merkur.de	1849
www.tagesspiegel.de	1540
www.bild.de	1442

Gesamtanzahl an URLS: 55622

Alice Weidel

www.welt.de	9507
www.focus.de	5287
www.merkur.de	4563
www.bild.de	4387
www.stern.de	3398
www.spiegel.de	3246
www.zeit.de	2781
www.huffingtonpost.de	2769
www.faz.net	2647
web.de	2248

Gesamtanzahl an URLS: 56290

Angela Merkel

www.welt.de	12 072
www.spiegel.de	11 327
www.focus.de	6491
www.zeit.de	4748
www.bild.de	3275
www.faz.net	2266
www.t-online.de	1679
www.sueddeutsche.de	1514
www.stern.de	1493
web.de	1304

Gesamtanzahl an URLS: 52983

Cem Özdemir

www.bild.de	5125
www.welt.de	4677
www.t-online.de	4110
www.faz.net	3013
www.bento.de	2343
www.rp-online.de	2280
www.spiegel.de	2261
www.n-tv.de	2130
www.focus.de	2085
www.general-anzeiger-bonn.de	1989

Gesamtanzahl an URLS: 55211

Christian Lindner

www.welt.de	9679
www.focus.de	3687
www.huffingtonpost.de	2934
www.faz.net	2841
www.rp-online.de	2454
www.bild.de	2380
www.bento.de	2105
www.n-tv.de	2025
www.merkur.de	1799
www.tagesspiegel.de	1764

Gesamtanzahl an URLS: 54296

Dietmar Bartsch

www.huffingtonpost.de	3120
www.welt.de	2742
www.deutschlandfunk.de	2623
www.focus.de	2272
www.mdr.de	2109
www.t-online.de	1899
www.stimme.de	1829
www.bento.de	1636
www.theeuropean.de	1439
www.svz.de	1433

Gesamtanzahl an URLS: 45251

Katrin Göring-Eckardt

www.stern.de	6977
www.n-tv.de	4438
www.epochtimes.de	3581
www.focus.de	3433
www.spiegel.de	2419
www.tagesspiegel.de	1731
www.rp-online.de	1670
jungefreiheit.de	1661
www.huffingtonpost.de	1641
www.bild.de	1552

Gesamtanzahl an URLS: 49993

Martin Schulz

www.welt.de	9790
www.spiegel.de	8330
www.faz.net	5582
www.focus.de	4909
www.zeit.de	4705
www.bild.de	4041
www.t-online.de	3610
www.n-tv.de	2868
www.sueddeutsche.de	2851
www.stern.de	985

Gesamtanzahl an URLS: 55865

Sahra Wagenknecht

www.welt.de	5673
www.bild.de	3070
www.focus.de	2523
www.heute.de	2495
www.rp-online.de	2446
www.abendzeitung-muenchen.de	2242
www.tagesspiegel.de	2216
www.faz.net	2094
www.ksta.de	2031
www.bento.de	1702

Gesamtanzahl an URLS: 52970

Abbildung 11: Häufigste Top-Level-Domains bei den Schlagzeilen zur Google-Suche nach den Personen

AfD

www.welt.de	39 244
www.tagesspiegel.de	31 952
www.focus.de	31 283
www.zeit.de	28 041
www.faz.net	26 921
www.merkur.de	24 214
www.spiegel.de	19 365
www.stern.de	17 351
www.sueddeutsche.de	10 603
www.rp-online.de	10 174

Gesamtanzahl an URLS: 363612

Bündnis90/Die Grünen

www.t-online.de	1287
www.focus.de	1192
www.schwaebische.de	1091
www.deutschlandfunk.de	897
www.suedkurier.de	855
www.berliner-kurier.de	835
www.chiemgau24.de	786
www.tagesspiegel.de	783
www.mdr.de	714
www.ndr.de	638

Gesamtanzahl an URLS: 23185

CDU

www.rp-online.de	36 789
www.faz.net	24 870
www.welt.de	24 054
www.tagesspiegel.de	17 276
www.focus.de	17 100
www.spiegel.de	11 671
www.stern.de	10 037
www.zeit.de	9867
www.deutschlandfunk.de	7957
www.vorwaerts.de	6459

Gesamtanzahl an URLS: 359399

CSU

www.sueddeutsche.de	52 591
www.merkur.de	42 533
www.spiegel.de	25 436
www.welt.de	23 973
www.tz.de	13 920
www.zeit.de	13 857
www.br.de	9774
www.infranken.de	9107
www.augsburger-allgemeine.de	8886
www.nordbayern.de	8856

Gesamtanzahl an URLS: 363838

Die Linke

www.welt.de	28 278
www.tagesspiegel.de	23 126
www.zeit.de	17 214
www.rp-online.de	14 283
www.faz.net	13 515
www.focus.de	12 980
www.neues-deutschland.de	11 786
www.ndr.de	9977
www.merkur.de	9883
www.deutschlandfunk.de	9570

Gesamtanzahl an URLS: 368838

FDP

www.rp-online.de	39 548
www.faz.net	19 483
www.welt.de	18 982
www.liberale.de	15 398
www.tagesspiegel.de	14 918
www.zeit.de	14 175
www.stern.de	12 087
www.focus.de	10 460
www.spiegel.de	10 403
www.handelsblatt.com	9541

Gesamtanzahl an URLS: 359884

SPD

www.faz.net	44 276
www.welt.de	33 237
www.vorwaerts.de	22 408
www.rp-online.de	22 259
www.zeit.de	20 190
www.tagesspiegel.de	17 629
www.focus.de	17 471
www.sueddeutsche.de	15 332
www.spiegel.de	11 707
www.stern.de	9615

Gesamtanzahl an URLS: 361836

Abbildung 12: Häufigste Top-Level-Domains bei den Suchen auf Google News nach den Parteien

Alexander Gauland

www.welt.de	46 917
www.faz.net	31 792
www.zeit.de	22 606
www.tagesspiegel.de	17 657
www.spiegel.de	16 475
www.merkur.de	15 815
www.focus.de	14 806
www.fr.de	13 275
www.deutschlandfunk.de	12 210
www.maz-online.de	7864

Gesamtanzahl an URLS: 366262

Alice Weidel

www.welt.de	30 800
www.zeit.de	20 899
www.stern.de	19 654
www.focus.de	19 546
www.rp-online.de	18 685
www.augsburger-allgemeine.de	17 197
www.huffingtonpost.de	16 907
www.spiegel.de	16 591
www.merkur.de	14 673
www.faz.net	13 994

Gesamtanzahl an URLS: 374593

Angela Merkel

www.faz.net	42 455
www.spiegel.de	40 131
www.welt.de	39 180
www.zeit.de	34 219
www.focus.de	24 392
www.stern.de	21 726
www.rp-online.de	18 585
www.tagesspiegel.de	14 626
www.merkur.de	7845
www.berliner-zeitung.de	6912

Gesamtanzahl an URLS: 363632

Cem Özdemir

www.tagesspiegel.de	23 854
www.rp-online.de	19 903
www.faz.net	18 671
www.bild.de	18 661
www.welt.de	12 954
www.badische-zeitung.de	12 874
www.focus.de	12 764
www.n-tv.de	12 228
www.t-online.de	12 177
www.shz.de	11 210

Gesamtanzahl an URLS: 367973

Christian Lindner

www.stern.de	30 111
www.rp-online.de	20 675
www.welt.de	19 215
www.focus.de	18 426
www.liberale.de	17 059
www.tagesspiegel.de	16 142
www.faz.net	15 998
www.spiegel.de	13 521
www.deutschlandfunk.de	13 233
www.bild.de	13 042

Gesamtanzahl an URLS: 372799

Dietmar Bartsch

www.welt.de	33 577
www.stimme.de	19 552
www.suedkurier.de	16 805
www.tagesspiegel.de	14 681
www.huffingtonpost.de	13 815
www.focus.de	10 496
www.svz.de	9896
www.all-in.de	8985
www.mdr.de	8377
www.shz.de	7845

Gesamtanzahl an URLS: 383192

Katrin Göring-Eckardt

www.stern.de	22 451
www.stimme.de	19 573
www.spiegel.de	19 337
www.welt.de	16 966
www.shz.de	16 505
www.rp-online.de	15 592
www.moz.de	12 620
www.n-tv.de	12 031
www.zeit.de	11 751
www.bild.de	11 293

Gesamtanzahl an URLS: 381740

Martin Schulz

www.faz.net	59 881
www.zeit.de	27 884
www.focus.de	26 605
www.welt.de	26 377
www.spiegel.de	21 546
www.stern.de	16 281
www.sueddeutsche.de	14 461
www.rp-online.de	11 419
www.tagesspiegel.de	10 810
www.merkur.de	10 409

Gesamtanzahl an URLS: 364193

Sahra Wagenknecht

www.welt.de	41 944
www.tagesspiegel.de	25 158
www.waz-online.de	21 906
www.ksta.de	20 084
www.focus.de	18 403
www.noz.de	15 237
www.rp-online.de	10 443
www.bild.de	10 217
www.lokalkompass.de	9559
www.huffingtonpost.de	9148

Gesamtanzahl an URLS: 384376

Abbildung 13: Häufigste Top-Level-Domains bei den Suchen auf Google News nach den Personen

3.4 Owned Content, Social Media und Medienangebote

Die Suchergebnislisten enthalten verschiedene Kategorien von Ergebnissen, wie ein Blick auf Abbildung 14 am Beispiel einer Suche nach "Cem Özdemir" zeigt.

Abbildung 14: Aufteilung der organischen Suchergebnisse zum Suchbegriff „Cem Özdemir" in drei mögliche Kategorien: Persönliche Webseiten, die man noch in „Owned Content" und Social Media unterteilen kann, editierbare Webseiten wie Wikipedia und Nachrichten. Daneben kommen noch relativ häufig Nachrichtenaggregatoren wie Freemail-Portale vor (t-online, web.de), die aber bei dieser Suche nicht unter den Ergebnissen waren.

Eine interessante Frage bei der Bewertung der Diversität der von Suchmaschinen ausgelieferten Ergebnisse besteht darin, ihre Verteilung auf verschiedene Angebotskategorien zu betrachten und dabei insbesondere die Frage zu stellen, inwieweit die Personen oder Parteien deren Inhalte selbst editieren können. Zu Letzteren gehören offensichtlich die eigenen Webseiten der Kandidaten und Parteien. Hierzu wurde jede einzelne URL der Datenbasis manuell entsprechend ihrer Top-Level-Domain einer von 7 verschiedene Kategorien zugeordnet und URLs der Kategorie „Medien" zusätzlich differenzierter kategorisiert[12]. Eine genaue Beschreibung der einzelnen Kategorien ist in Appendix A nachzulesen.

Tabelle 3: Verteilung aller Top-Level-Domains der Google-Suche auf die Angebotskategorien.

Owned Content	3947
Medien	406
Öffentlich finanziert	303
Sonstige	71
Social Media	12
Wikipedia	10
Freemail Portale	5

Es ist mehr als erstaunlich, dass von den 4754 im Untersuchungszeitraum ausgelieferten Top Level-Domains bei der Google Suche der weit überwiegende Anteil von 83 % der Editierbarkeit der Parteien oder Personen unterliegt („Owned Content"). Der weit überwiegende Teil des Owned Content besteht aus den Seiten der Ortsverbänden der jeweiligen Parteien (siehe Tabelle 4). Dabei sind einige Parteien weitaus aktiver als andere: Während die Suchergebnisse in Summe für die AfD nur 6 klar erkennbare eigene Top Level-Domains liefern, sind es bei der SPD mehr als 1000, und bei der CDU knapp 950. Neben den Webseiten der Parteien und Personen und den ca. 400 Mediendomains sticht auch die relative hohe Anzahl an öffentlich finanzierten Domains heraus (gut 300 Domains). Diese repräsentieren haupt-

[12] Wir danken der Bayerischen Landeszentrale für Neue Medien für diese mühevolle Arbeit.

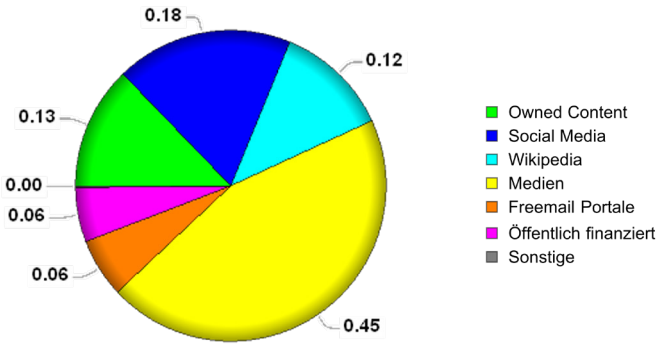
sächlich Städte und Gemeinden; dazu gehören aber auch die Webseite der Bundeszentrale für politische Bildung und die Webseiten des Bundestages.

Tabelle 4: Anzahl der Top-Level-Domains, die einer Partei zugeordnet werden können. Diese repräsentieren meistens einen Ortsverband, die Partei selbst oder einzelne Parteiangehörige.

Partei	Anzahl einer Partei zuordenbaren Top-Level-Domains
SPD	1043
CDU	944
FDP	687
Bündnis90/Die Grünen	675
Die Linke	465
CSU	96
AfD	6

In Abbildung 15 ist die Verteilung der organischen Suchergebnisse (ohne Schlagzeilen) über den gesamten Untersuchungszeitraum aufgetragen. Hierbei sind deutliche Unterschiede beim Anteil von URLs der Kategorie „Owned Content“ und „Medien“ zwischen Personen und Parteien festzustellen. Der Anteil an Links, die auf eine Wikipedia-Seite verweisen, liegt über 10 %, also etwas über einem Link, da es Nutzer gibt, bei denen sowohl der deutsche als auch der englische Wikipedia-Eintrag ausgeliefert wurde. Durch diese Kategorisierung ist es nun auch möglich, zwischen dem gesamten Anteil an URLs, die eine Partei oder Person editieren kann (Owned Content, Social Media und grundsätzlich auch Wikipedia) und den Anteil, den sie nicht editieren kann (Medien, Freemail-Portale, Öffentlich finanziert und Sonstige), zu unterscheiden.

Personen:



Parteien:

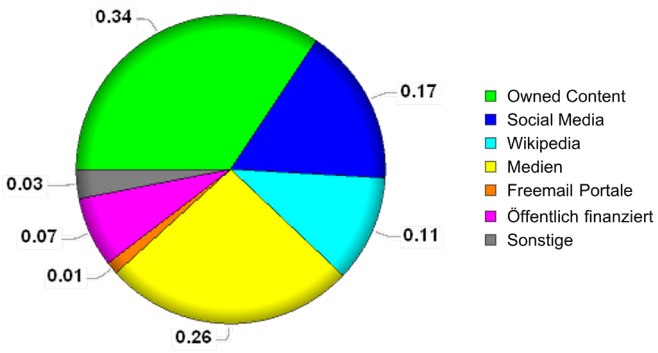


Abbildung 15: Verteilung der organischen Suchergebnisse auf die verschiedenen Kategorien in %, links für die Personen, rechts für die Parteien. Die 9-10 organischen Suchergebnisse eines Nutzers bei einer Suche nach einer Person enthielten also z. B. im Schnitt zwei URLs aus der Kategorie „Social Media“ und 4-5 Links aus der Kategorie „Medien“.

Wenn man den Anteil der jeweiligen Kategorien für die einzelnen Parteien zusammenfasst (siehe Tabelle 5), zeigen sich weitere Anzeichen für eine unterschiedliche mediale Präsenz im digitalen Raum. Während Bündnis90/Die Grünen mit knapp 60 % und Die Linken sowie die FDP mit um die 50 % einen sehr hohen Anteil an Owned Content, also Links auf eigene Domains von Parteien, Ortsverbänden oder Politikern erreicht haben, schaffen es CDU und SPD lediglich auf 30 %. Auch hier ist natürlich nicht klar, was Henne und was Ei ist: Eine gute Search Engine Optimization (SEO) sorgt für eine gute Sichtbarkeit der Inhalte in allen Suchmaschinen, aber eine gute Sichtbarkeit in Suchmaschinen kann auch für eine hohe Popularität der Inhalte sorgen, die wiederum für eine gute Sichtbarkeit in den Suchmaschinen sorgt. Es ist hier also nicht bewertbar, was der Grund dafür ist, dass für manche Parteien der eigene Inhalt besser sichtbar ist als für andere. Nach allen hier getätigten Untersuchungen halten wir es für am wahrscheinlichsten, dass die Parteien gut sichtbar sind, die sowohl auf der Regionalisierungsebene als auch auf der Nachrichtenebene für genügend aktuelle Inhalte sorgen. Tabelle 5 zeigt dieselbe Analyse für Suchbegriffe, die Personen repräsentieren. Hier fällt auf, dass insbesondere bei Gauland, der keine eigenen Social-Media-Accounts hatte, insgesamt weniger direkt oder im Prinzip editierbare Domains ausgeliefert werden – bei den meisten anderen Personen werden ca. zwei Social-Media-Accounts ausgeliefert. Für diejenigen mit einer Bundestagsmandatur wird oft auch eine öffentlich finanzierte Domain angegeben (z. B. www.bundestag.de, www.bundeskanzlerin.de), die dementsprechend für Gauland, Weidel, Schulz und Lindner fehlen.

Zuletzt zeigt die Abbildung 16 noch die Verteilung der Kategorien bei den ausgelieferten Schlagzeilen in der Google-Suche. Es ließen sich keine signifikanten Unterschiede sowohl auf Suchbegriffsebene als auch aggregiert nach den Parteien und Personen feststellen. Der überwiegende Teil von 85 % kommt stets aus dem Bereich der Printmedien mit einer Online-Seite, die den Printteil wiedergibt. Medienangebote, deren einziger Verbreitungsweg das Internet ist, werden zu 9 % in den Schlagzeilen angezeigt und lediglich 3 % entfallen jeweils auf öffentlich-rechtliche Kanäle und die Onlineableger klassischer deutscher privater Fernsehanbieter.

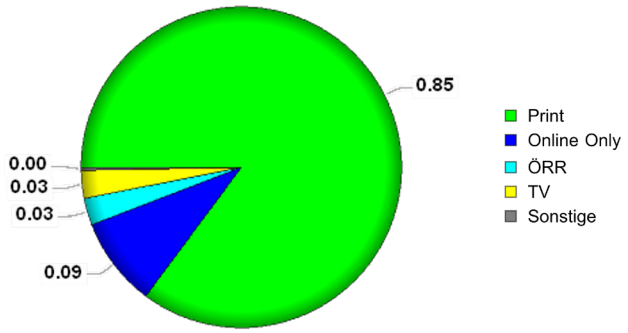


Abbildung 16: Anteil der von Google ausgelieferten Medien in den Topstorys im Untersuchungszeitraum bei der Google-Suche über alle Suchbegriffe.

Tabelle 5: Verteilung der organischen Sucherergebnisse auf die verschiedenen Kategorien, für die Suchen nach den Parteien. Die Kreisdiagramme repräsentieren den Anteil aller Webseiten, die (im Prinzip) durch die Person oder Partei editierbar sind in rot (Owned Content, Social Media und Wikipedia) und den Anteil der Webseiten, auf denen ein solcher Eingriff nicht möglich ist (Freemail Portale, Medien, öffentlich finanziert und Sonstige), in blau.

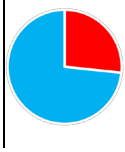
	AfD	Bündnis90/ Die Grünen	CDU	CSU	Die Linke	FDP	SPD
Owned Content	0,12	0,59	0,32	0,20	0,46	0,41	0,30
Social Media	0,11	0,16	0,19	0,20	0,21	0,21	0,10
Wikipedia	0,04	0,11	0,12	0,12	0,15	0,13	0,12
Medien	0,60	0,01	0,25	0,42	0,07	0,19	0,29
Freemail Portale	0,03	0,00	0,02	0,01	0,00	0,01	0,02
Öffentlich finanziert	0,09	0,05	0,09	0,04	0,07	0,06	0,10
Sonstige	0,01	0,08	0,00	0,00	0,04	0,00	0,07
Anteil der (im Prinzip) selbst editierbaren Webseiten (rot)/ ohne einen solchen Zugriff (blau)							

Tabelle 6: Verteilung der organischen Sucherergebnisse auf die verschiedenen Kategorien, für die Suchen nach den Personen. Die Kreisdiagramme repräsentieren den Anteil aller Webseiten, die (im Prinzip) durch die Person oder Partei editierbar sind in rot (Owned Content, Social Media und Wikipedia) und den Anteil der Webseiten, auf denen ein solcher Eingriff nicht möglich ist (Freemail Portale, Medien, öffentlich finanziert und Sonstige), in blau.

	Alexander Gauland	Alice Weidel	Angela Merkel	Cem Özdemir	Christian Lindner
Owned Content	0,08	0,11	0,11	0,16	0,14
Social Media	0,01	0,22	0,15	0,23	0,23
Wikipedia	0,12	0,12	0,11	0,12	0,12
Medien	0,71	0,52	0,46	0,27	0,47
Freemail Portale	0,08	0,03	0,04	0,13	0,02
Öffentlich finanziert	0,00	0,00	0,13	0,09	0,01
Sonstige	0,00	0,00	0,00	0,00	0,01
Anteil der (im Prinzip) selbst editierbaren Webseiten (rot)/ ohne einen solchen Zugriff (blau)					

	Dietmar Bartsch	Katrin Göring-Eckardt	Martin Schulz	Sahra Wagenknecht
Owned Content	0,16	0,12	0,11	0,16
Social Media	0,22	0,20	0,22	0,17
Wikipedia	0,12	0,12	0,12	0,12
Medien	0,29	0,37	0,50	0,46
Freemail Portale	0,10	0,08	0,04	0,03
Öffentlich finanziert	0,11	0,10	0,00	0,06
Sonstige	0,00	0,01	0,00	0,00
Anteil der (im Prinzip) selbst editierbaren Webseiten (rot)/ ohne einen solchen Zugriff (blau)				

4 Raum für Personalisierung / Regionalisierung

Die Datenspende Bundestagswahl 2017 ist ein *proof-of-concept*, um nachzuweisen, dass die Gesellschaft mit Hilfe von Black-Box-Analysen in der Lage ist, algorithmische Entscheidungssysteme wie Suchmaschinen – aber prinzipiell auch andere wie Produktempfehlungssysteme oder soziale Medien – auf mögliche Gefahrenstellen zu untersuchen. Eine viel diskutierte Gefahrenstelle der letzten Jahre war die algorithmenbasierte Bildung und Verhärtung von Filterblasen auf Grundlage von personalisierten Such- und Empfehlungsfiltern. Die Frage danach, ob und inwieweit Suchmaschinennutzer unterschiedliche Nachrichten und Informationen zu wichtigen politischen Institutionen und Personen bekommen, stand daher im Vordergrund der Analyse. Dass wir eine Suchmaschine und nicht die viel anfälligeren sozialen Medien wie Facebook untersucht haben, lag vor allen Dingen am leichteren Zugang zu den Ergebnissen.
Unter Personalisierung der Suchergebnisse verstehen wir, wie bereits ausgeführt, eine auf den Nutzer hin zugeschnittene Aufbereitung oder Auswahl von Suchergebnissen, die er oder sie noch nicht angeklickt hat. Die Vorschläge werden aufgrund von Profilen und vom Klickverhalten anderer, dem Suchenden als ähnlich empfundenen Nutzern abhängig gemacht.
Die Personalisierung selbst kann in der getroffenen Auswahl, aber auch in einer unterschiedlichen Reihenfolge in der Präsentation der gefundenen Ergebnisse zum Ausdruck kommen. In den Erläuterungen zur Nutzung des Google-Dienstes heißt es dazu:

> *"Unsere automatisierten Systeme analysieren Ihre Inhalte (einschließlich E-Mails), um Ihnen für Sie relevante Produktfunktionen wie personalisierte Suchergebnisse, personalisierte Werbung sowie Spam- und Malwareerkennung bereitzustellen."*[13].

Wie groß war also die Personalisierung für die Datenspenderinnen und Datenspender vor der Bundestagswahl 2017?

4.1 Anteile gleicher Suchergebnisse

Das Buch von Eli Pariser ging auf die vielen möglichen Gefahren von Filterblasen ein und äußerte die Sorge, dass jede Person in ihrer eigenen Filterblase gefangen sei ("being isolated in a web of one"[14]), die zudem eher "information desserts" als nahrhafte Informationen ("information vegetables") enthalte[15]. Daher ist eine erste interessante Frage, welcher Anteil von Paaren von Datenspendern exakt dieselben Links angezeigt bekommen und welcher Anteil dazu sogar noch diese Links in exakt derselben Reihenfolge bekommt.

Bei den Ergebnissen der Google-Suche (ohne Schlagzeilen) ergeben sich die in Tabelle 7 zusammengefassten Resultate für den Anteil gleicher Ergebnislisten und gleicher Ergebnislisten in exakt derselben Reihenfolge. Tatsächlich liegen diese Anteile bei den Parteien im Wesentlichen im niedrigen einstelligen Prozentbereich – abgesehen von der AfD, bei der 16 % der Datenspenderpaare dieselben Ergebnisse enthalten. Es ist wichtig zu betonen, dass dies **nicht** bedeutet, dass 16 % der Ergebnislisten exakt gleich sind. Dazu ein Beispiel: Wenn von 100 Personen zwei Gruppen von Personen mit je 5 und 6 Mitgliedern exakt dieselben Ergebnisse bekommen und alle anderen mit niemandem die exakt gleichen Links teilen, dann gibt es einmal 10 Paare von Nutzern und einmal 15 Paare von Nutzern, die jeweils die gleichen Links sehen, von insgesamt 4.950 Paaren. Das entspricht 0.51 % aller Paare und damit fast dem Ergebnis für die CDU.

[13] https://policies.google.com/privacy?gl=de#infouse, abgerufen am 15.5.2018
[14] Aus Eli Parisers TED Talk: https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles, Minute 8.
[15] Aus Eli Parisers TED Talk: https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles, Minute 5-6.

Tabelle 7: Anteil (in Prozent) gleicher Suchergebnisse ohne Berücksichtigung der Sortierung, getrennt nach Suchbegriffen für Parteien und Personen (2. Spalte). In Spalte 3 befindet sich der Anteil (in Prozent) gleicher Suchergebnisse in der exakt gleichen Sortierung.

Partei	Anteil gleicher Ergebnisse	Anteil gleicher Ergebnisse mit Sortierung
AfD	16.5001	1.83138
Bündnis90/Die Grünen	0.57114	0.214303
CDU	0.570018	0.149749
CSU	1.70591	0.353519
Die Linke	3.46704	1.24836
FDP	1.34052	0.343161
SPD	2.88426	0.599181

Person	Anteil gleicher Ergebnisse	Anteil gleicher Ergebnisse mit Sortierung
Alexander Gauland	8.83055	3.22469
Alice Weidel	29.6601	8.45194
Angela Merkel	26.6007	9.05307
Cem Özdemir	32.0659	6.76066
Christian Lindner	31.0667	11.2727
Dietmar Bartsch	34.5123	3.50788
Katrin Göring-Eckardt	45.0472	10.4351
Martin Schulz	26.6496	5.50891
Sahra Wagenknecht	44.7985	21.1722

Wirklich erstaunlich ist der enorm hohe Anteil von paarweise gleichen Suchergebnissen bei den Personen, der – außer bei Alexander Gauland – bei durchschnittlich mindestens einem Viertel und bei manchen fast bei 50% liegt. D. h., hätte man zwei beliebige Datenspender gebeten, zu demselben Zeitpunkt eine Suche nach einer der Personen zu machen, wären bei einem Viertel bis knapp der Hälfte dieser Paare exakt dieselben Links herausgekommen – und bei ca. 5-10 % auch noch in derselben Reihenfolge. Wenn hier Filterblasen vorlägen, wären zumindest einige davon recht geräumig und würden jeweils größere Gruppen von Personen gemeinsam betreffen.

Eine manuelle Inspektion der Ergebnislisten von Alexander Gauland lässt vermuten, dass insbesondere die fehlenden Social-Media-Accounts zu diesem Ergebnis führen: Diese werden durch Nachrichten ersetzt, die anscheinend diverser ausgeliefert werden. Ein Rückblick auf Abbildung 9 unterstützt diese Hypothese. Die Abbildung zeigt, dass die Politikerinnen und Politiker unter ihren Top 10 Top-Level-Domains unterschiedlich viele „Owned Content“ und „editierbare“ Domains aufweisen. Tabelle 8 stellt den prozentualen Anteil gleicher Suchergebnislisten der Anzahl von "Owned Content"- und "Social Media"-Top-Level-Domains in den Top 10 der Suchergebnis-Top-Level-Domains gegenüber. Es wird sichtbar, dass mehr von diesen Top-Level-Domains unter den Top 10 mit einer höheren Übereinstimmung der Suchergebnislisten korrelieren.

Tabelle 8: Gegenüberstellung der Art von Top-Level-Domains der Politikerinnen und Politiker unter ihren Top 10 Top-Level-Domains (Abbildung 9) und der prozentuale Anteil gleiche Ergebnislisten.

Suchbegriff	Anzahl „Owned Content" + „Social Media" unter Top 10 Top-Level-Domains in der organischen Suche	Prozentualer Anteil gleicher Ergebnislisten
Alexander Gauland	3	9%
Angela Merkel	4	27%
Martin Schulz	4	27%
Alice Weidel	5	29%
Christian Lindner	4	31%
Cem Özdemir	4	32%
Dietmar Bartsch	4	35%
Sahra Wagenknecht	5	45%
Katrin Göring-Eckardt	4	45%

In jedem Fall bleibt festzuhalten, dass auch bei Personen, die dieselben Links erhalten, die Reihenfolge wahrscheinlich anders ist, da der Anteil derjenigen mit exakt derselben Reihenfolge von denen mit denselben Ergebnissen in den meisten Fällen unter 50 % liegt.

4.2 Durchschnittliche Anzahl gemeinsamer Links

Nur weil zwei Personen nicht exakt dieselben Links bekommen, heißt das natürlich nicht, dass sie völlig unterschiedliche Links bekommen. Daher zeigen die Tabelle 9 und Tabelle 10 jeweils die durchschnittliche Anzahl gemeinsamer Links für alle Paare von Datenspendern, die zu demselben Suchzeitpunkt denselben Suchbegriff nachschlagen. Die Zahl liegt für die Personen zwischen 7,2 und 8,1 gemeinsamen Links. Dies muss noch zur durchschnittlichen Länge der Suchergebnislisten in Beziehung gesetzt werden. Da diese nur knapp über 9 beträgt, sind also im Mittel nur 1 bis knapp 2 Links pro zufällig gezogenem Paar von Datenspendern unterschiedlich. Dies ist der „Raum für Personalisierung", also die Anzahl an Suchergebnissen, die nun mit Links unterschiedlichen politischen Inhalts gefüllt sein könnten. In jedem Fall ist aber damit – im Durchschnitt! – die Informationslage über die Personen im Wesentlichen gleich. Wir stellen also durchschnittlich eine sehr hohe Überlappung der jeweils erhaltenen Nachrichten und Informationen fest.

Tabelle 9: Die mittlere Ergebnislistenlänge ohne Schlagzeilen (3. Spalte) für die Personen gegenübergestellt der durchschnittlichen Anzahl an URLs, die sich eine Ergebnisliste mit allen anderen Ergebnislisten am selben Suchzeitpunkt teilt (2. Spalte). Daraus lässt sich durch Subtraktion der durchschnittliche Raum für eine Personalisierung berechnen (4. Spalte).

Person	gemeinsame Links	mittlere Länge der Suchergebnisliste	Raum für Personalisierung
Alexander Gauland	7.21006	9.06237	1.8523
Alice Weidel	7.5889	9.06172	1.47282
Angela Merkel	7.88147	9.02418	1.1427
Cem Özdemir	7.76558	9.08891	1.32333
Christian Lindner	7.49458	9.08401	1.58943
Dietmar Bartsch	8.1262	9.24993	1.12373
Katrin Göring-Eckardt	8.11471	9.18402	1.0693
Martin Schulz	7.49725	9.02739	1.53014
Sahra Wagenknecht	7.96757	9.11063	1.14307

Bei den Parteien sieht das Ergebnis etwas anders aus. Hier haben zwei beliebige Datenspender, die beispielsweise zu demselben Zeitpunkt nach „Die Linke“ suchen, im Durchschnitt nur 4,7 gemeinsame Links, bei der „AfD“ bis zu 7 gemeinsame Links. Vergleicht man dies wieder mit den durchschnittlichen Längen der Suchergebnislisten, die zwischen 8 und 9 schwanken[16], bleiben für die Parteien im Durchschnitt zweier Datenspender zwischen 2 (AfD), 3 (Bündnis90/Die Grünen, Die Linke, FDP) und fast 4 (CDU, CSU) **nicht** geteilter Links übrig.

Tabelle 10: Die mittlere Ergebnislistenlänge ohne Schlagzeilen (3. Spalte) für die Parteien gegenübergestellt der durchschnittlichen Anzahl an URLs, die sich eine Ergebnisliste mit allen anderen Ergebnislisten am selben Suchzeitpunkt teilt (2. Spalte). Daraus lässt sich durch Subtraktion der durchschnittliche Raum für eine Personalisierung berechnen (4. Spalte).

Partei	gemeinsame Links	mittlere Länge der Suchergebnisliste	Raum für Personalisierung
AfD	6.99774	9.03863	2.04089
Bündnis90/Die Grünen	6.03871	9.15723	3.11852
CDU	5.39071	8.94959	3.55888
CSU	5.31047	8.99783	3.68736
Die Linke	4.70699	8.00472	3.29773
FDP	5.1696	8.35014	3.18054
SPD	6.28109	9.03634	2.75524

Die Abbildung 17 und Abbildung 18 zeigen, dass die Durchschnitte im Wesentlichen aus leicht linksschiefen Verteilungen stammen. Das liegt daran, dass jedes Paar höchstens 8 oder 9 Links teilen kann, je nach durchschnittlicher Länge der Listen. Nach rechts ist also diese Grenze näher als nach links. Nur bei der AfD sieht man eine zweigipflige Verteilung, also z.B. 10 % der Datenspenderpaare, die sich nur 2 Links teilen. Dieser Befund wird im Weiteren noch betrachtet (Abschnitt 5).

Dieselbe Untersuchung haben wir auch noch auf den Suchergebnislisten von Google News durchgeführt, deren Ergebnis wir im folgenden Abschnitt diskutieren.

[16] Die kurzen organischen Suchergebnislisten sind ein Effekt von angezeigten Schlagzeilen (dann enthält die organische Suchergebnisliste meistens 9 Einträge) und kann weiterhin vermindert werden durch die Anzeige von Löschhinweisen oder ähnlichem. Insbesondere bei „Die Linke“ scheinen solche Effekte die Suchergebnislisten verkürzt zu haben. Da die entsprechenden Informationen auf der Suchergebnisseite aber nicht mitgespeichert wurden, können wir hier keine abschließende Bewertung vornehmen.

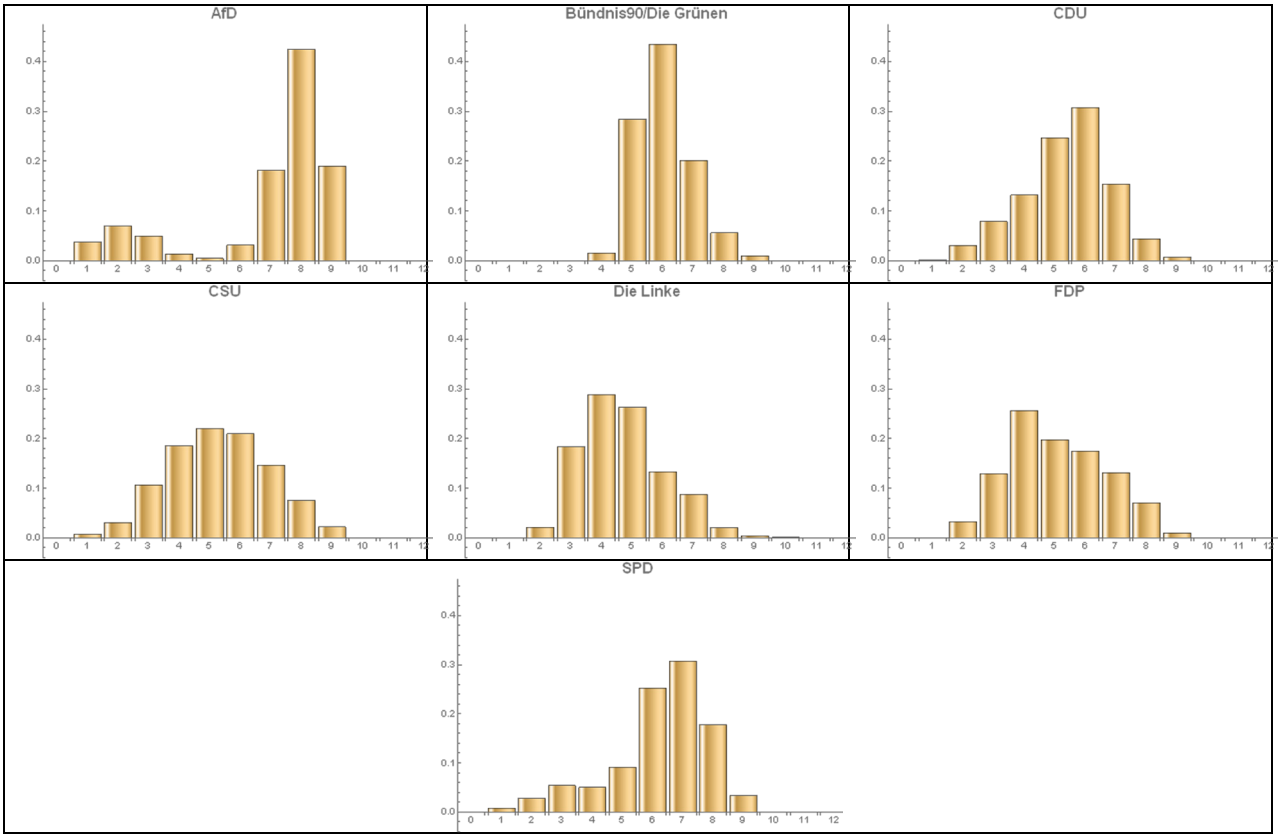


Abbildung 17: Verteilung der Anzahl gleicher URLs im paarweisen Vergleich bei den organischen Suchergebnissen zu den Parteien.

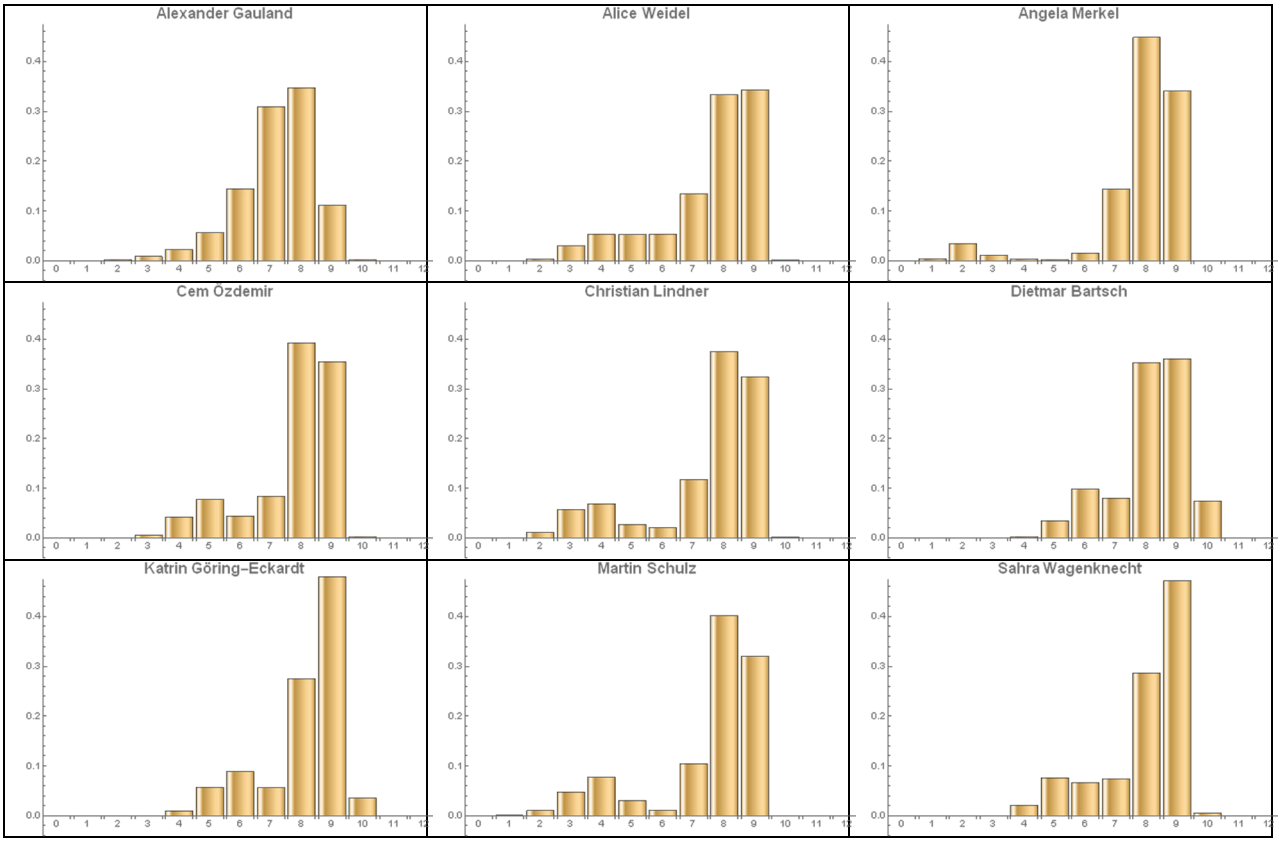


Abbildung 18: Verteilung der Anzahl gleicher URLs im paarweisen Vergleich bei den organischen Suchergebnissen zu den Personen.

4.3 Raum für Personalisierung bei Google News

Die Suchergebnislisten bei Google News enthalten fast immer 20 Einträge, dementsprechend liegen die mittleren Längen gerundet auch alle bei 20. Von diesen sind bei den Personen im Durchschnitt 17,3 bis 18,3 geteilt. Prozentual ist also der Anteil geteilter Links sogar höher als bei der Google-Suche, es bleiben 2 bis 3 Links mit möglicherweise personalisierten Nachrichten übrig.

Person	durchschnittliche Anzahl gemeinsamer Links	mittlere Länge der Ergebnislisten	durchschnittlicher Raum für Personalisierung
Alexander Gauland	18.0928	19.9667	1.87385
Alice Weidel	17.2749	19.8514	2.57648
Angela Merkel	18.3288	19.991	1.66218
Cem Özdemir	18.1502	19.9789	1.82867
Christian Lindner	18.078	19.9387	1.86071
Dietmar Bartsch	18.0043	19.8279	1.82363
Katrin Göring-Eckardt	17.2947	19.7863	2.49161
Martin Schulz	17.8846	19.9697	2.08504
Sahra Wagenknecht	18.2229	19.944	1.7211

Abbildung 19: Die mittlere Ergebnislistenlänge auf Google News (3. Spalte) für die Personen gegenübergestellt mit der durchschnittlichen Anzahl an URLs, die sich eine Ergebnisliste mit allen anderen Ergebnislisten am selben Suchzeitpunkt teilt (2. Spalte). Daraus lässt sich durch Subtraktion der durchschnittliche Raum für eine Personalisierung berechnen (4. Spalte).

Auch bei den Parteien ist der Raum für personalisierte Nachrichten (im Durchschnitt) eher gering und liegt bei 1,3 bis 2,1 für alle bis auf „Bündnis90/Die Grünen“, bei denen bis zu 3 Links von 20 personalisiert sein könnten. Es ist wichtig zu beachten, dass hier immer nur Paare von Personen betrachtet werden. Das, was sich A und B **nicht teilen**, könnte sich aber B mit C teilen. Damit müssen auch diese nicht geteilten Links nicht notwendigerweise nur einer Person angezeigt werden, sondern können wieder kleinere Teilgruppen betreffen.

Partei	durchschnittliche Anzahl gemeinsamer Links	mittlere Länge der Ergebnislisten	durchschnittlicher Raum für Personalisierung
AfD	18.3519	19.9802	1.62838
Bündnis90/Die Grünen	14.6315	17.7313	3.09981
CDU	18.3747	19.7574	1.38266
CSU	18.6942	19.9999	1.30568
Die Linke	17.4219	19.5355	2.1136
FDP	18.1947	19.7516	1.55692
SPD	18.4332	19.8903	1.45712

Abbildung 20: Die mittlere Ergebnislistenlänge auf Google News (3. Spalte) für die Parteien gegenübergestellt mit der durchschnittlichen Anzahl an URLs, die sich eine Ergebnisliste mit allen anderen Ergebnislisten am selben Suchzeitpunkt teilt (2. Spalte). Daraus lässt sich durch Subtraktion der durchschnittliche Raum für eine Personalisierung berechnen (4. Spalte).

5 Regionalisierung

Durch die hohe Anzahl an „Owned Content“, der hauptsächlich den vielen Webseiten regionaler Ortsverbände oder den Webseiten lokaler Politiker der Parteien geschuldet ist, kommt es bei einer Regionalisierung der Suchergebnisse zwangsläufig zu vielen nicht-geteilten Links. Daher haben wir den jeweils einer Partei zugehörigen „Owned Content“ nochmals in regional/nicht-regional kategorisiert. Dabei sind wir eher konservativ vorgegangen und haben nur solche Domains als regional markiert, die in der URL klar einen Städtebezug aufwiesen. Für jedes Paar von Suchergebnislisten zu Parteien haben wir nun zuerst diese regionalen URLs entfernt und dann die durchschnittliche Anzahl nicht geteilter Links berechnet. Dies stellt also eine genauere Annäherung an die mögliche Anzahl personalisierter Links dar, weil die regionalen Links explizit ausgeklammert werden. Tabelle 11 zeigt das Ergebnis dieser Verfeinerung: Während die Datenspender sich bei den Grünen im Durchschnitt bei 3 Links unterscheiden, ist nur einer davon nicht-regionaler Natur (und könnte damit personalisiert sein). Ähnlich ist es bei SPD, FDP, CDU und den Linken, die jeweils von (gut) 3 nicht geteilten Links auf unter 2 nicht-regionale Links fallen.

Tabelle 11: Verfeinerung der durchschnittlichen Anzahl nicht geteilter Links unter Berücksichtigung der Regionalisierung. In der linken Spalte steht die durchschnittliche Anzahl nicht geteilter Links bezüglich der gesamten Suchergebnisliste, in der rechten bezüglich der Suchergebnisliste nach dem Löschen der klar regionalisierten Webseiten (der jeweiligen Ortsverbände oder einzelner, lokaler Politikerinnen und Politiker).

	Nicht geteilte Links	Nicht geteilte Links (ohne regionalisierte Links)
Bündnis90/Die Grünen	3.11852	1.02717
SPD	2.75524	1.572
FDP	3.18054	1.70385
CDU	3.55888	1.84831
Die Linken	3.29773	1.8818
AfD	2.04089	2.0244
CSU	3.68736	3.36667

Bei der AfD zeigt sich im Wesentlichen kein Effekt – das verwundert aber auch nicht, da nur sechs Top-Level-Domains überhaupt angezeigt werden und keine Ortsverbandwebseiten gefunden wurden. Von den sechs Domains sind vier regionaler Natur (afd.berlin, afd.nrw, afd-bw.de, afd-fraktion-brandenburg.de). Interessanterweise ergibt sich auch bei der CSU nur wenig Änderung. Ein Blick auf die Deutschlandkarte mit den ungefähren Standorten unserer Datenspender zeigt aber, dass Bayern weniger Datenspender aufweist, die zudem in den ersten vier Wochen des Untersuchungszeitraums noch in den Sommerferien waren. Da die meisten Ortsverbände in Bayern liegen, wurden also kaum regionalisierte Suchergebnisse ausgespielt. Eine manuelle Inspektion der Suchergebnislisten zeigt, dass relativ viele Nachrichten ausgeliefert werden, die sich weniger stark überlappen. Hier könnte also eine tiefergehende, inhaltliche Inspektion sinnvoll sein. Nichtsdestotrotz überlappen sich die Suchergebnislisten im Durchschnitt zu deutlich mehr als der Hälfte, so dass auch hier eine gemeinsame Informationsbasis gelegt wird.

6 Dynamik

Eine wichtige Teilfrage des Projektes betrifft die Dynamik, mit der sich Suchergebnisse bei der normalen Google-Suche verändern. Eine mögliche Hypothese wäre hier, dass sich die von Google ausgelieferten Suchergebnisse in ihrer Ähnlichkeit oder den Anteilen verschiedener Medien zur Wahl hin verändern. In der Reihenfolge der bisherigen Analysen haben wir hier zunächst die Dynamik des jeweiligen Anteils von Schlagzeilen als am prominentesten platzierte Ergebnisse betrachtet (siehe Abschnitt 6.1). Als zweites folgt eine Analyse der Anzahl ausgerollter Top-Level-Domains pro Suchzeitpunkt (siehe Abschnitt 6.2), gefolgt von einer Betrachtung der jeweiligen Anteile an selbst editierbaren URLs an den Suchanfragen, also "Owned Content", Social Media und (mit Abstrichen) Wikipedia (siehe Abschnitt 6.3).

6.1 Dynamik des Vorkommens von Schlagzeilen

Ein erster interessanter Punkt bei der Frage, ob sich die Suchergebnisse zur Wahl hin verändert haben, war, ob und wann Schlagzeilen mit angezeigt werden und ob sich der Anteil der Suchergebnislisten mit Schlagzeilen beispielsweise gegen Ende des Wahlkampfes erhöht hat. Hierbei könnten Effekte wie eine höhere Nachrichtendichte dazu führen, dass Google kurz vor der Wahl beispielsweise mehr Suchergebnislisten mit Schlagzeilen ausrollt als 5 Wochen davor.

Die Analyse des Vorkommens von Schlagzeilen hat allerdings ergeben, dass über alle Suchbegriffe und Suchzeitpunkte hinweg der Anteil kaum schwankt. Die Anteile bewegen sich stets in einem Korridor um die 90% mit Ausschlägen nach oben und unten von maximal 10 % (siehe Abbildung 21). Erst bei Betrachtung der einzelnen Suchbegriffe haben sich stärkere Ausschläge feststellen lassen.

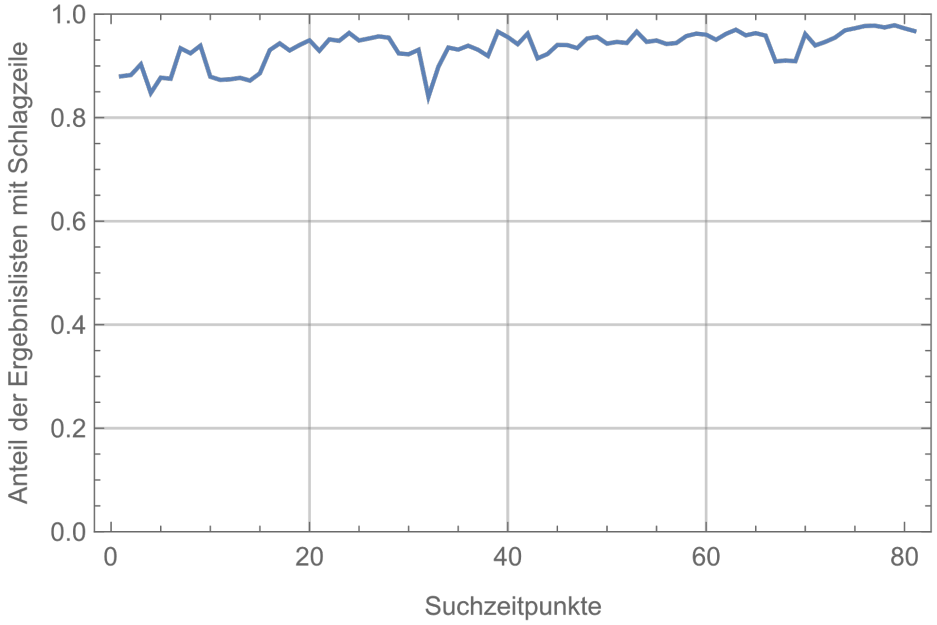


Abbildung 21: Anteil aller Ergebnislisten zu einem Suchzeitpunkt, bei denen mindestens eine Schlagzeile ausgeliefert wurde.

So gibt es deutliche Ausreißer bei Katrin Göring-Eckhardt und Dietmar Bartsch (siehe Abbildung 22). Sehr stabil zeigen sich dagegen die Anteile der Suchergebnislisten mit Schlagzeilen für die Suchbegriffe CSU, SPD, Angela Merkel, Martin Schulz sowie die AfD, wobei deren Schlagzeilenanteil zu Beginn des Untersuchungszeitraums (Zeitpunkt 6 bis 22) noch Schwankungen unterliegt (siehe Abbildung 23). Das vereinzelte, vollständige Fehlen von Schlagzeilen (siehe Abbildung 22) könnte darauf hinweisen, dass von Google je nach aktueller Nachrichtenlage entschieden wird, ob eine Schlagzeile ausgeliefert wird – vollständig erklären lassen sich diese Schwankungen jedoch von außen nicht. Es ist ohne weitere Informationen insbesondere nicht nachvollziehbar, wieso Google an manche Studienteilnehmer zu demselben Suchzeitpunkt und derselben Anfrage Schlagzeilen ausliefert und an manche nicht. Gerade die sehr stabilen Werte bei einigen Suchbegriffen (CSU, SPD, Angela Merkel und Martin Schulz) sprechen deutlich gegen eine Browsereinstellung dieser Nutzer, die eine solche Schlagzeilenanzeige in irgendeiner Form unterdrücken würde, denn dann hätten diese Nutzer auch bei den anderen Suchbegriffen (wie CSU, SPD, Angela Merkel) keine Schlagzeilen erhalten. Diese Suchbegriffe zeigen aber viele Suchzeitpunkte auf, bei denen alle Nutzer Schlagzeilen bekamen.

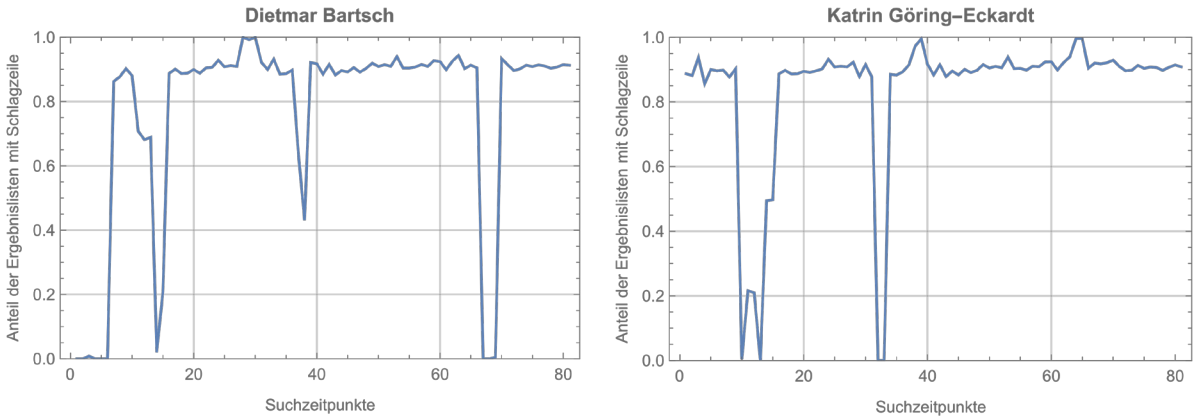


Abbildung 22: Anteil der Ergebnislisten mit Schlagzeilen bei den Suchbegriffen „Dietmar Bartsch“ und „Katrin Göring Eckhardt“ zeigen deutliche Abweichungen von allen anderen.

Deutlich wird diese Inkonsistenz auch bei der Betrachtung einzelner Nutzer, so werden zum Beispiel bei demselben Nutzer und demselben Suchbegriff manchmal Schlagzeilen angezeigt und manchmal nicht (siehe Abbildung 24). Insgesamt ist aber der Anteil der Suchergebnislisten mit Schlagzeilen relativ konstant und weist insbesondere keine zeitlichen Muster hinsichtlich der Bundestagswahl auf.

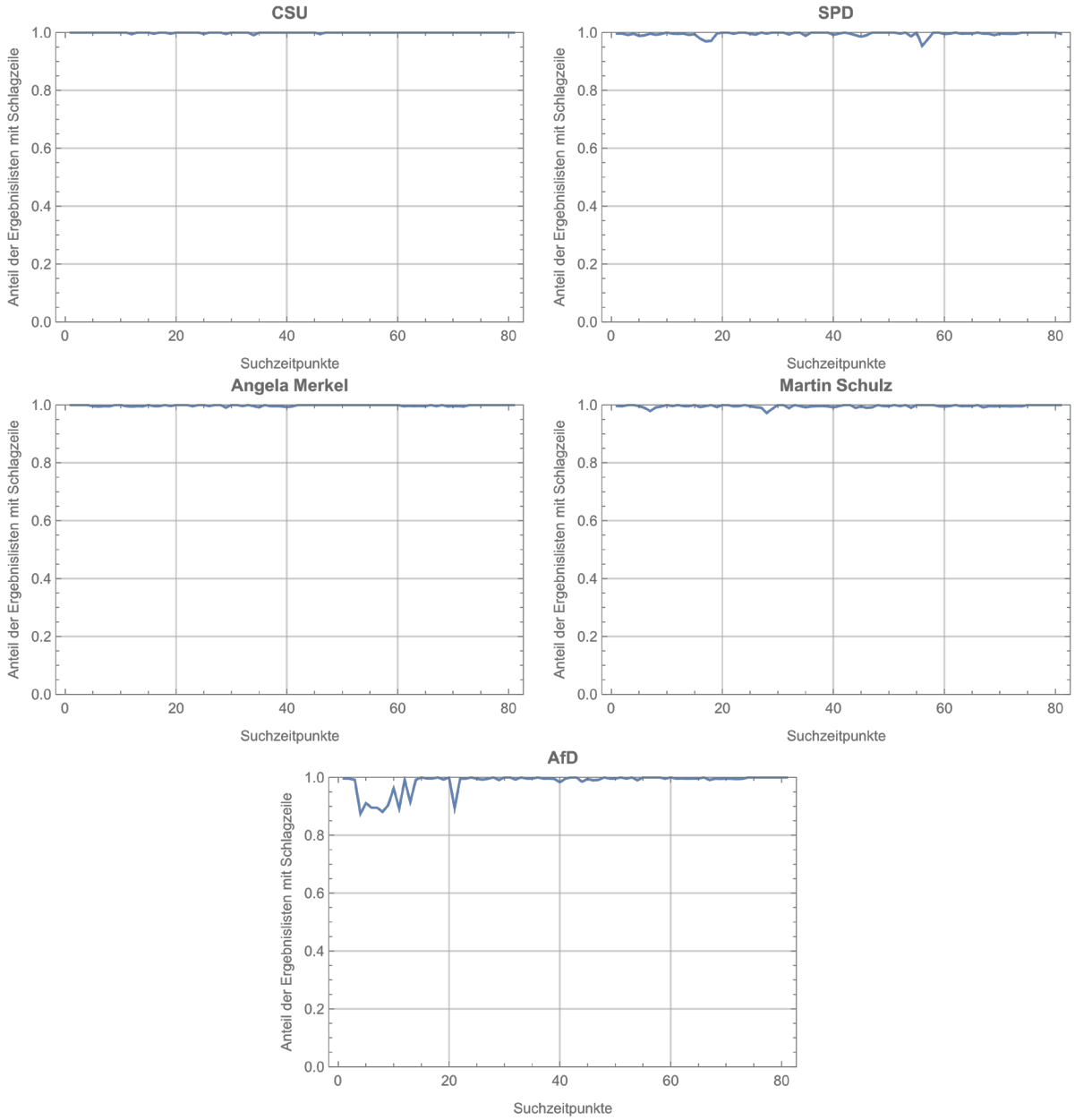


Abbildung 23: Anteil aller Ergebnislisten, die Schlagzeilen enthalten, bei den Suchbegriffen CSU, SPD, Angela Merkel, Martin Schulz und AfD.

00:01 Uhr	04:01 Uhr
http://www.stimme.de/themen/wahlen/btw17/wahlchecks/Wahlcheck-mit-Katrin-Goering-Eckardt;art140707,3901145	
http://www.rp-online.de/politik/deutschland/katrin-goering-eckardt-sieht-tiefe-kluft-zwischen-gruenen-und-fdp-aid-1.7027881	
http://www.stimme.de/themen/wahlen/btw17/wahlchecks/Ohne-Klimaschutz-ohne-die-Gruenen;art140707,3901145	
http://www.goering-eckardt.de/	http://www.goering-eckardt.de/
https://de.wikipedia.org/wiki/Katrin_G%C3%B6ring-Eckardt	https://de.wikipedia.org/wiki/Katrin_G%C3%B6ring-Eckardt
https://de-de.facebook.com/GoeringEckardt/	https://de-de.facebook.com/GoeringEckardt/
http://www.spiegel.de/thema/katrin_goering_eckardt/	http://www.spiegel.de/thema/katrin_goering_eckardt/
https://www.bundestag.de/abgeordnete18/biografien/G/goering_eckardt_katrin/258372	https://www.bundestag.de/abgeordnete18/biografien/G/goering_eckardt_katrin/258372
http://www.focus.de/politik/deutschland/gruene-spitzenkandidatin-goering-eckardt-hat-sich-von-ihrem-mann-getrennt-und-spricht-ueber-ihre-neue-liebe_id_6967147.html	http://www.focus.de/politik/deutschland/gruene-spitzenkandidatin-goering-eckardt-hat-sich-von-ihrem-mann-getrennt-und-spricht-ueber-ihre-neue-liebe_id_6967147.html

Abbildung 24: Suchergebnisse eines Nutzers oder einer Nutzerin vom 25.08.2017 zu den genannten Zeiten und zum Suchbegriff „Katrin Göring-Eckardt". Demselben Nutzer werden zu unterschiedlichen, konsekutiven Suchzeitpunkten einmal Schlagzeilen angezeigt (grau hinterlegt) und einmal nicht.

6.2 Dynamik der Anzahl unterschiedlicher Top-Level-Domains

Bei der Aufbereitung der Anzahl unterschiedlicher Top-Level-Domains, die mindestens ein Datenspender oder eine -spenderin zu einem Suchzeitpunkt als Ergebnis bekam, bestätigen sich einerseits bereits erfasste Trends (siehe Abbildung 25Abbildung 26), dass nämlich Suchergebnislisten für Parteien grundsätzlich mehr verschiedene Top-Level-Domains aufweisen als Ergebnislisten für Personen. Andererseits schwankt diese Zahl auch sehr stark von Partei zu Partei: Während nur ca. 22 Top-Level-Domains die Ergebnisse über die AfD an den meisten Suchzeitpunkten bestimmen, sind es bei Bündnis90/Die Grünen um die 180 (mit großen Ausschlägen). Im Wesentlichen ist hier noch einmal die Anzahl der Top-Level-Domains der Parteien widergespiegelt und damit auch das schwankende Ausmaß der Regionalisierung sichtbar.

Bei den Personen liegt die Anzahl jeweils zwischen 17 und 18, was ebenfalls zeigt, dass die Anzahl geteilter Links (also auch Top-Level-Domains) bei den Personen sehr hoch ist.

Auch wenn die Schwankungen über den Suchzeitraum zum Teil sehr stark sind, ergeben sich keine zeitlich auffälligen Muster in der Anzahl unterschiedlicher Top-Level-Domains.

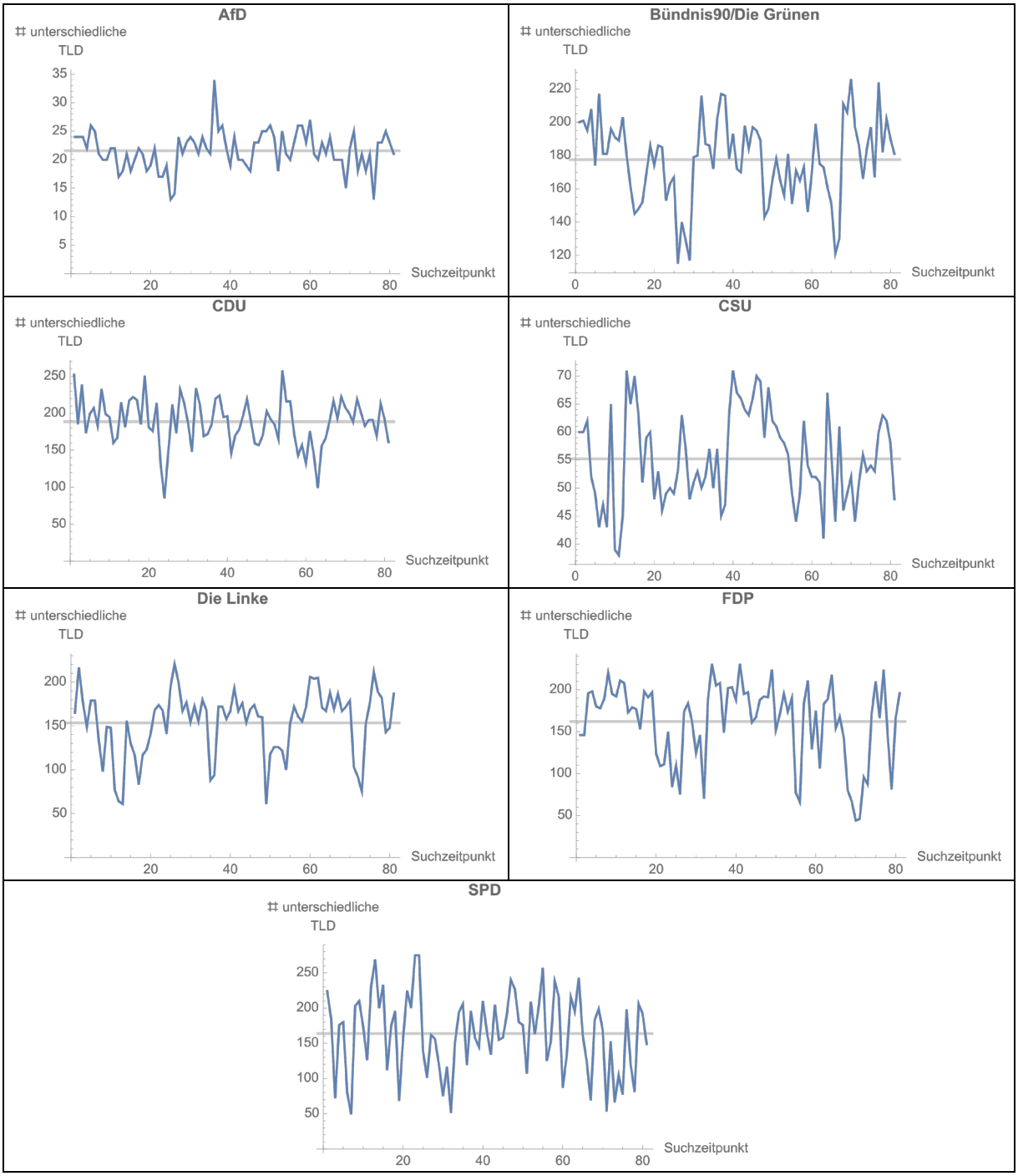


Abbildung 25: Anzahl unterschiedlicher Top-Level-Domains über die 81 Suchzeitpunkte bei der Google-Suche nach den Parteien.

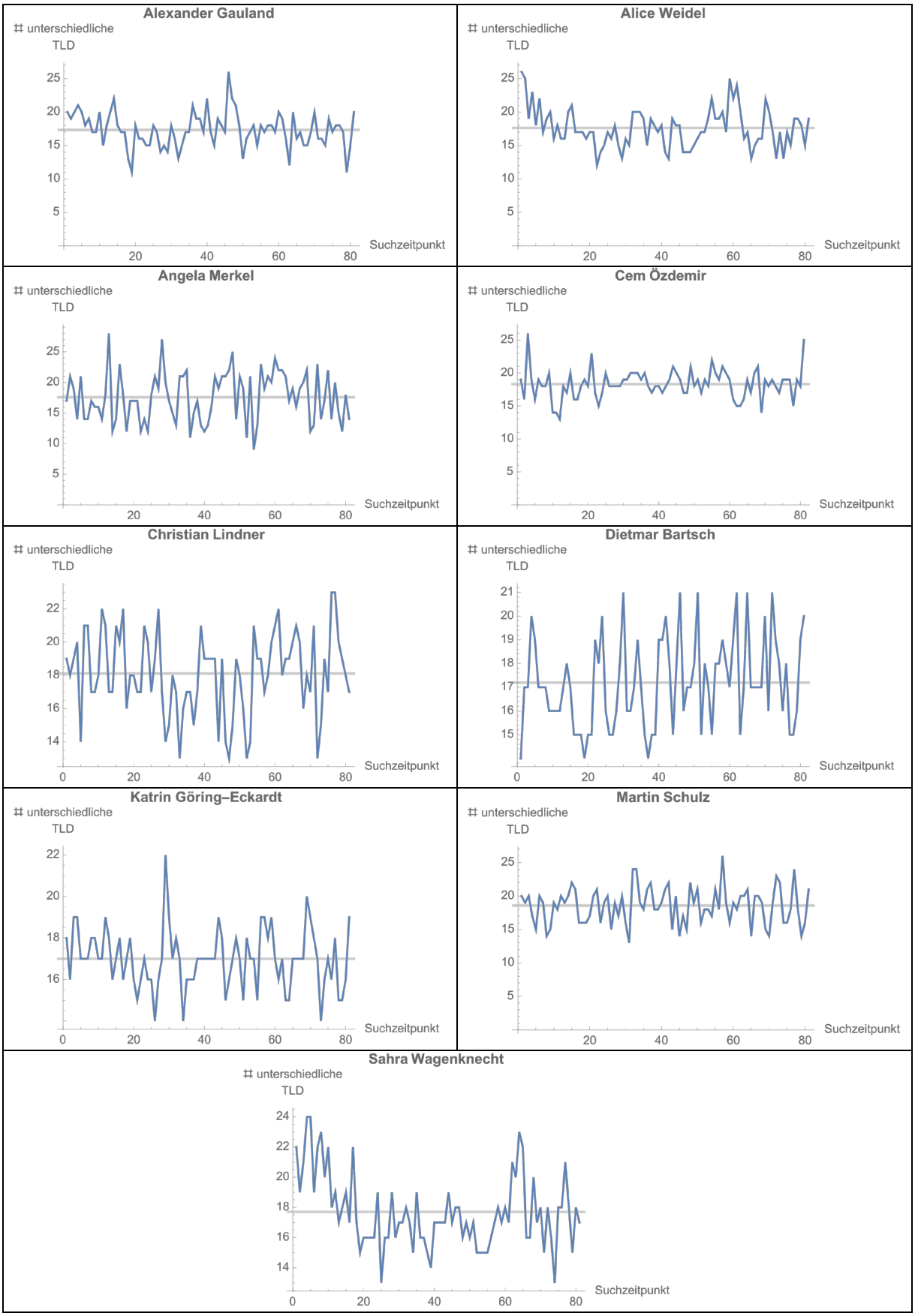


Abbildung 26: Anzahl unterschiedlicher Top-Level-Domains über die 81 Suchzeitpunkte bei der Google-Suche nach den Personen.

6.3 Dynamik der Anteile editierbarer Beiträge

Verändert sich über die Zeit der Anteil der URLs, die auf Top-Level-Domains liegen, die von den Parteien oder Personen (im Prinzip) editierbar sind? Auch diese Frage haben wir für alle Suchzeitpunkte im Untersuchungszeitraum untersucht. Es ergibt sich ein kleiner Abfall für alle Parteien (außer den Grünen) für die Suchzeitpunkte direkt vor der Wahl. Dazu ist zu bemerken, dass die letzten 6 Suchzeitpunkte vom Wahlwochenende stammen. Vermutlich wird dies durch einen erhöhten Nachrichtenanteil erzeugt, der aber erst in den letzten Tagen vor der Wahl zum Tragen kommt.

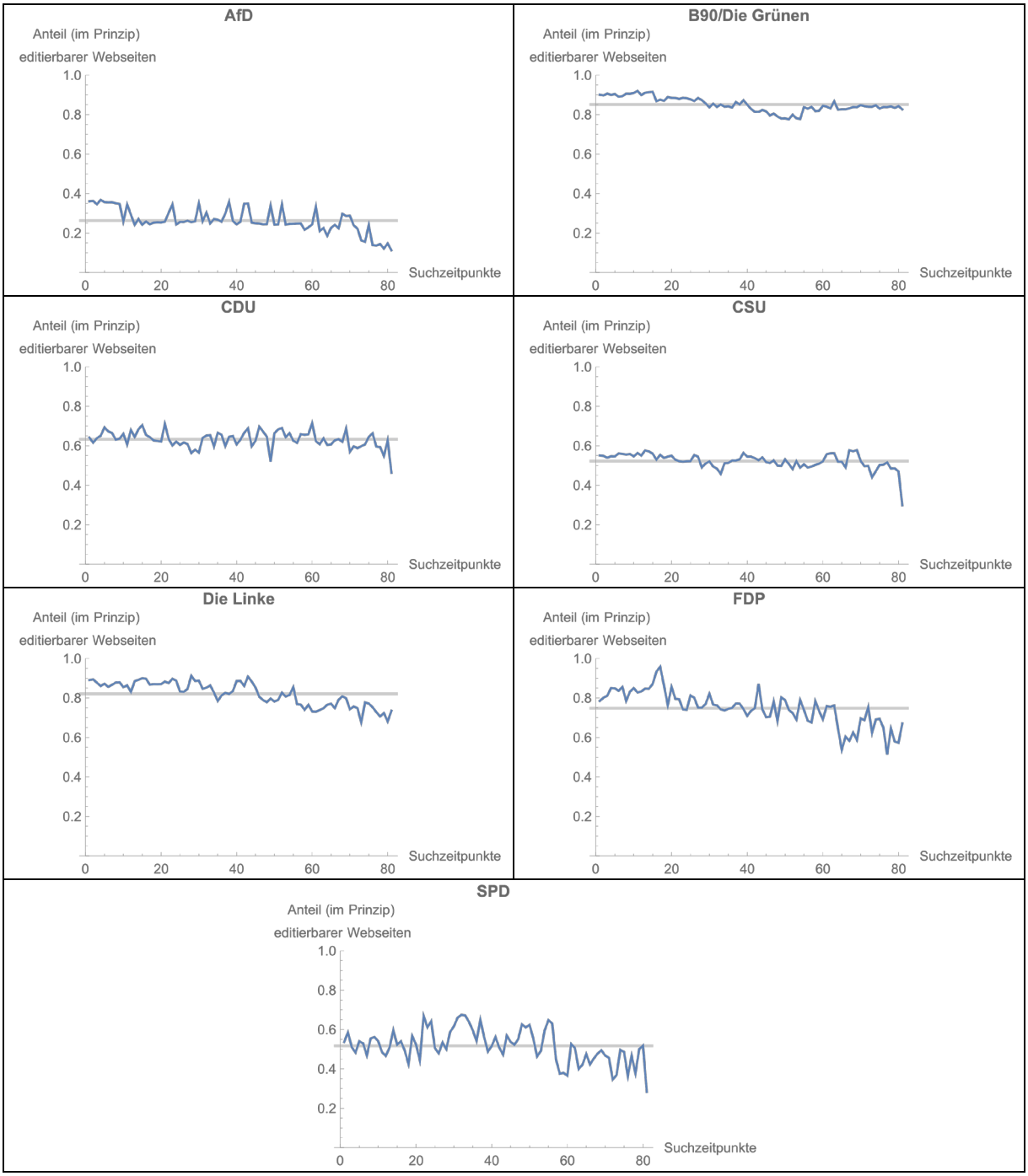


Abbildung 27: Anteil der für die jeweilige Partei (im Prinzip) editierbaren Webseiten an allen Ergebnislinks der jeweiligen Partei. Hierbei ist der Anteil für jeden der 81 Suchzeitpunkte aufgetragen und der Mittelwert in grau eingezeichnet.

Auch bei den meisten Personen ist dieser kurzfristige Rückgang an selbsteditierbaren Webseiten unter den Suchergebnissen zu sehen, der bei manchen schon eine Woche vor der Wahl beginnt (Gauland, Lindner, Özdemir und Wagenknecht). Das Ergebnis wurde nicht weiterverfolgt, da es in diesem Bericht um die wesentlichen Fragen nach Personalisierung, Regionalisierung und Filterblasen geht. Es könnte aber für Medienwissenschaftlerinnen und -wissenschaftler als eine interessante Ausgangsbeobachtung herausstellen, der mit den öffentlich verfügbaren Daten der Datenspende weiter nachgegangen werden kann.

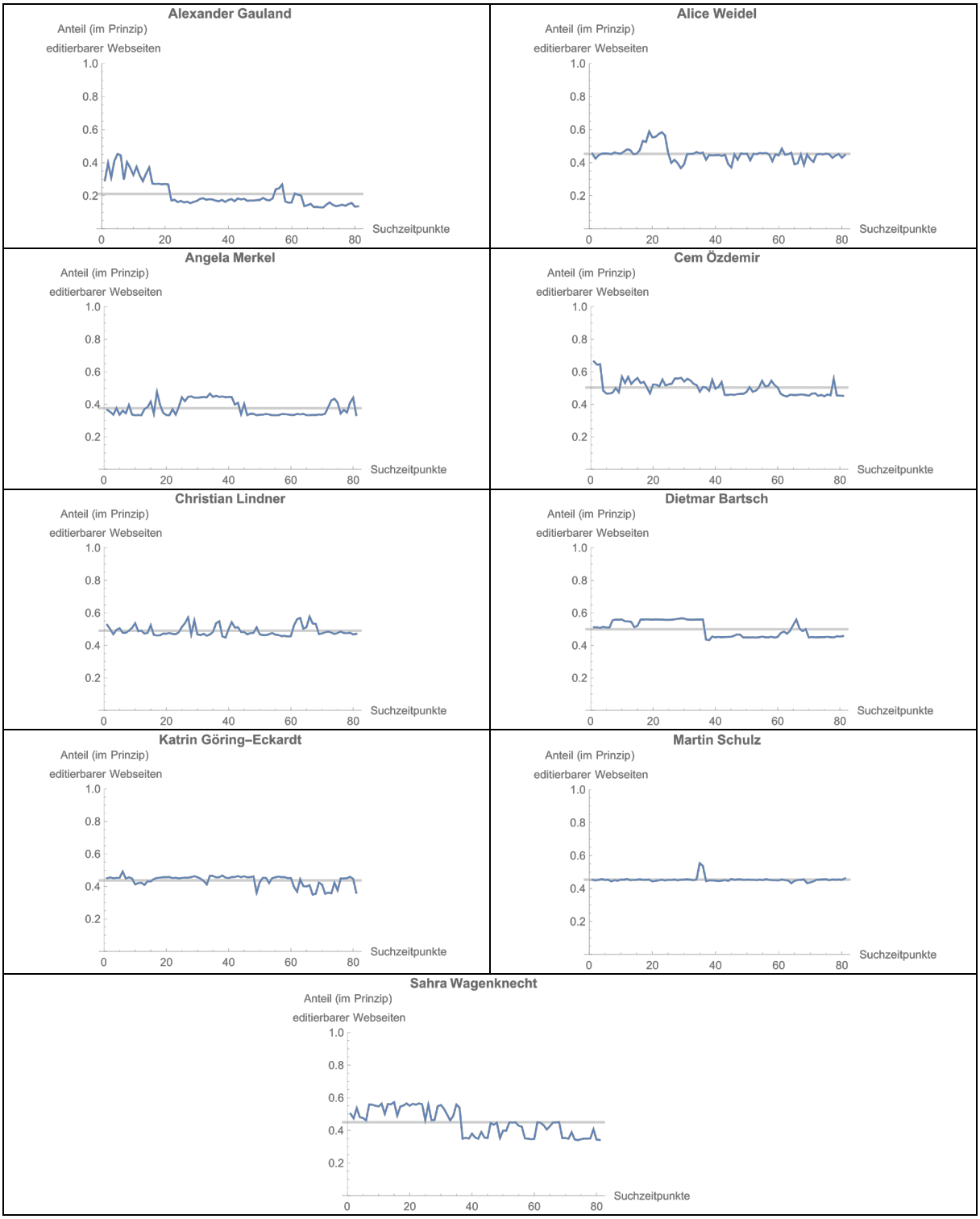


Abbildung 28: Anteil der für die jeweilige Person (im Prinzip) editierbaren Webseiten an der Suche nach der jeweiligen Person. Hierbei ist der Anteil für jeden der 81 Suchzeitpunkte aufgetragen und der Mittelwert in Grau eingezeichnet.

7 Suchergebnislisten mit geringer Ähnlichkeit zu allen anderen

Auch wenn die meisten Suchergebnislisten sich im Durchschnitt stark ähneln, können sich hinter Mittelwerten auch immer Paare von Suchergebnislisten verbergen, die sich kaum überlappen. Schon im ersten Zwischenbericht wurde erwähnt, dass eine manuelle Inspektion der Suchergebnislisten immer wieder einzelne Listen ans Licht bringt, die sich deutlich von den anderen unterscheiden[17]. Es wurde festgestellt, dass diese Listen oft Links auf Webseiten in anderen Sprachen (hauptsächlich englisch) enthalten. Es ist aber schwierig, diese Cluster zweifelsfrei und automatisch zu identifizieren. Ein Teil der Schwierigkeit liegt darin, dass man dem Link selbst nicht unbedingt ansieht, in welcher Sprache die dahinterliegende Webseite verfasst ist, beispielsweise enden manche Webseiten auf ".com", die dann trotzdem deutschsprachige Seiten anzeigen, wie "handelsblatt.com". Social Media enden auf ".com" (twitter.com, facebook.com) und die angezeigte Sprache hängt davon ab, ob der Nutzer in seinem Account eingeloggt ist. Bei der Wikipedia ist die Sprache in den ersten zwei Buchstaben der URL codiert, so weisen "de.wikipedia.org" auf deutsche und "en.wikipedia.org" auf englische Inhalte hin. Außerdem ist nicht klar, ob alle abweichenden Cluster mindestens eine fremdsprachige Webseite enthalten müssen.

Wir haben auch einen zweiten Ansatz verfolgt, der ebenfalls der manuellen Nachbearbeitung bedarf: In einem ersten Schritt wurden Links entfernt, die für einen Suchbegriff zu einem genannten Suchzeitpunkt von mindestens 70 % aller Nutzerinnen und Nutzer gesehen wurden, da ihre Popularität gegen eine Personalisierung auf diesen Links spricht. In einem zweiten Schritt haben wir die Listen mit den verbliebenen Links in Cluster zusammengefasst, die untereinander wieder ähnliche Ergebnisse zeigen. Für diese Gruppen kann man dann berechnen, wie viele gemeinsame Einträge die ursprünglichen Listen der Gruppe mit allen Nutzern **außerhalb** der Gruppe haben. Ein Beispiel: Wir finden 5 Personen, die nur noch 4 Einträge auf ihren Listen behalten, wenn alle populären Links für den Suchbegriff „Die Linke“ daraus entfernt werden, aber alle 5 Personen behalten paarweise mindestens 3 gemeinsame Links – sie bilden ein Cluster. Für diese 5 Personen sehen wir uns jetzt wieder jeweils die gesamte Suchergebnisliste an und berechnen die durchschnittliche Anzahl gemeinsamer Links mit allen anderen Nutzern außerhalb des Clusters. Wenn diese im Durchschnitt unter 3,5 fällt, scheint eine sehr unterschiedliche Gruppe gefunden worden zu sein. Aber bei einer manuellen Inspektion der so identifizierten Cluster mussten wir feststellen, dass insbesondere bei den Suchbegriffen „Die Linke“ und „Bündnis90/Die Grünen“ hier hauptsächlich regionale Gruppen gefunden wurden, deren Suchergebnislisten nur speziell bezüglich ihrer Auswahl der vielen Ortsgruppen der genannten Parteien waren.

Zusammenfassend lässt sich also sagen, dass sich eine automatisierte Identifikation solcher Gruppen von Nutzern, die tatsächlich im Sinne der Filterblasentheorie inhaltlich abweichende Suchergebnisse bekommen, die nicht im Wesentlichen einer Regionalisierung der Ergebnisse entstammen, als sehr komplex erwiesen hat.

Aber ein dritter Zugang zeigte ein klares Muster, das tatsächlich auf eine stabile Filterblase hindeutet, die in weiteren Studien genauer untersucht werden müsste: Anhand der Zeitangaben, die den Schlagzeilen vorangestellt werden, kann man erkennen, welche Hauptsprache Google einem Nutzer zuteilt. In unserem Datensatz haben wir Zeitangaben auf Norwegisch, Englisch, Deutsch und Französisch gefunden. Diese Einschätzung hat eine Partitionierung der Suchergebnislisten zur Folge. In allen untersuchten Fällen sehen wir eine Unterteilung der Nutzer durch Google in „vermutlich Deutsche“ mit untereinander sehr ähnlichen Ergebnissen pro Suchzeitpunkt und Suchbegriff, und in „vermutlich Ausländer“, die untereinander jeweils außergewöhnliche Ergebnisse teilen, die dann mit Ergebnissen der vermuteten, bevorzugten Sprache gemischt werden.

[17] Abschnitt 4 im ersten Zwischenbericht (Krafft et al., 2017)

Es handelt sich dabei um manuelle Inspektionen, da es auch hier keine automatisierbare Unterteilung der Nutzer gibt in solche, die Google für deutschsprachig und solche, die es für nicht (primär) deutschsprachig hält. Manche Nutzer erhalten einmal deutschsprachige Zeitangaben, dann englischsprachige Zeitangaben, und viele der manuell erkannten Nutzer mit partiell fremdsprachigen Suchergebnissen erhalten keine Schlagzeilen und somit auch keine Zeitangaben.

Dieses Muster werden wir an ausgewählten Beispielen im Folgenden demonstrieren.

7.1 Beispiel: Suchergebnisse von einer französischen IP aus

24.8., 12h	24.8., 16h	25.8., 12h	25.8., 16h
http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930	http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930	http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930	
http://www.francetvinfo.fr/monde/europe/migrants/allemagne-des-refugies-syriens-nomment-leur-fille-angela-merkel-muhammed_2339261.html	http://www.francetvinfo.fr/monde/europe/migrants/allemagne-des-refugies-syriens-nomment-leur-fille-angela-merkel-muhammed_2339261.html	https://fr.sputniknews.com/societe/201708241032779081-angela-merkel-exposition-agriculture-virtuelle/	http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930
http://www.lemonde.fr/europe/article/2017/08/22/martin-schulz-attaque-angela-merkel-pour-decoller-dans-les-sondages_5175066_3214.html	http://www.lemonde.fr/europe/article/2017/08/22/martin-schulz-attaque-angela-merkel-pour-decoller-dans-les-sondages_5175066_3214.html	http://www.francetvinfo.fr/monde/europe/migrants/allemagne-des-refugies-syriens-nomment-leur-fille-angela-merkel-muhammed_2339261.html	https://fr.sputniknews.com/societe/201708241032779081-angela-merkel-exposition-agriculture-virtuelle/
https://fr.wikipedia.org/wiki/Angela_Merkel	https://fr.wikipedia.org/wiki/Angela_Merkel	https://fr.wikipedia.org/wiki/Angela_Merkel	https://fr.wikipedia.org/wiki/Angela_Merkel
https://de.wikipedia.org/wiki/Angela_Merkel	https://de.wikipedia.org/wiki/Angela_Merkel	https://de.wikipedia.org/wiki/Angela_Merkel	https://de.wikipedia.org/wiki/Angela_Merkel
http://www.gala.fr/stars_et_gotha/angela_merkel	http://www.gala.fr/stars_et_gotha/angela_merkel	http://www.gala.fr/stars_et_gotha/angela_merkel	http://www.gala.fr/stars_et_gotha/angela_merkel
http://www.france24.com/fr/tag/angela-merkel/	http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930	https://www.telestar.fr/people/angela-merkel-qui-est-le-mari-de-la-chanceliere-allemande-photos-198102	http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930
http://plus.lefigaro.fr/tag/angela-merkel	http://tempsreel.nouvelobs.com/tag/angela-merkel	http://www.parismatch.com/Actu/International/Angela-Merkel-au-pays-des-gamers-1331930	http://plus.lefigaro.fr/tag/angela-merkel
http://www.huffingtonpost.fr/news/angela-merkel/	http://plus.lefigaro.fr/tag/angela-merkel	http://tempsreel.nouvelobs.com/tag/angela-merkel	http://tempsreel.nouvelobs.com/tag/angela-merkel
http://tempsreel.nouvelobs.com/tag/angela-merkel	http://www.huffingtonpost.fr/news/angela-merkel/	http://plus.lefigaro.fr/tag/angela-merkel	http://www.huffingtonpost.fr/news/angela-merkel/
https://www.bundeskanzlerin.de/	http://www.lemonde.fr/europe/article/2017/08/22/martin-schulz-attaque-angela-merkel-pour-decoller-dans-les-sondages_5175066_3214.html	http://www.huffingtonpost.fr/news/angela-merkel/	http://www.france24.com/fr/tag/angela-merkel/
http://www.lemonde.fr/europe/article/2017/08/22/martin-schulz-attaque-angela-merkel-pour-decoller-dans-les-sondages_5175066_3214.html	http://www.france24.com/fr/tag/angela-merkel/	http://www.france24.com/fr/tag/angela-merkel/	https://www.bundeskanzlerin.de/

Abbildung 29: Suchergebnisse für den Suchbegriff "Angela Merkel" von einer Nutzerin oder einem Nutzer mit einer IP-Adresse in Frankreich, nahe Paris. Die Ergebnisse sind untergliedert in 3 Schlagzeilen und darauffolgend organische Suchergebnisse.

Dazu zeigen wir zuerst in Abbildung 29 die Suchergebnisse einer Nutzerin oder eines Nutzers in Frankreich. Bei den Suchergebnislisten mit Schlagzeilen kann man klar erkennen, dass Google von einem französischen Menschen ausgeht, da die Zeitangaben, welche von Google an jeden Nachrichtenbeitrag angefügt werden, auf Französisch sind (z. B., „Il y a 2 heurs“). Die Schlagzeilen selbst sind ebenfalls auf Französisch und betreffen hauptsächlich den deutschen Wahlkampf. Interessant ist, dass auch mehrere deutschsprachige Webseiten angezeigt werden. Insgesamt sehen die Ergebnisse zum Suchbegriff „Angela Merkel“ aber – für eine in Frankreich lebende Person – nach einer sinnvollen Auswahl aus.

Beim Suchbegriff „AfD“ kommen bei derselben Person neben Ergebnissen zur deutschen Partei Webseiten von französischen Initiativen mit demselben Kürzel zum Tragen, insbesondere die *L'Agence française de développement*, eine französische Vereinigung im Kampf gegen Diabetes und eine zur Unterstützung von Autisten und ihren Familien. Dasselbe Vorgehen nach regional und international bekannten, anderen Institutionen mit derselben Abkürzung zu suchen, sieht man auch beim Suchbegriff „CSU“. Diese kämpft mit den französischen Kürzeln für *Cultures et Sociétés Urbaines*, aber auch den amerikanischen Universitäten *Charles Sturt University* und der *Colorado State University* um Aufmerksamkeit.

Die Suche nach den deutschen Politikerinnen und Politikern läuft weitgehend fehlerfrei, in dem Sinne, dass tatsächlich fast immer Informationen über die gesuchte Person angezeigt werden: Im Wesentlichen werden die Social-Media-Profile gezeigt, daneben – je nach Bekanntheitsgrad der Politikerin oder des Politikers – ihre französische Wikipedia-Seite und französische Nachrichten; dies ist insbesondere bei Angela Merkel und Martin Schulz der Fall. Ansonsten werden deutsche, eigene Webseiten der Personen, ihre deutschen Social-Media-Accounts oder deutsche Nachrichten angezeigt. Bei Cem Özdemir kommt es zu zwei Verwechslungen: Einer der beiden angezeigten Twitter-Accounts gehört einem anderen Cem Özdemir[18] und auch bei dem Fußballspieler mit gleichem Namen, der auf dem Transfermarkt gehandelt wird[19], handelt es sich nicht um den Politiker der Grünen.

Der Verweis auf Links in der (vermuteten) Muttersprache des Nutzers oder der Nutzerin ist bei einer im Ausland lebenden Person ohne Frage sinnvoll und die vielen deutschsprachigen Links in den Suchergebnislisten der hier untersuchten Person sind vermutlich darauf zurückzuführen, dass es kaum französischsprachige Links für diese Suchbegriffe gibt. Wie genau hier die Abwägung getroffen wird, ob und wann einer Person im Ausland deutsche und Links auf Webseiten der entsprechenden Landessprache gezeigt werden, kann nicht weiter geklärt werden.

Es erweist sich als glücklicher Zufall, dass wir für den IP-Standort Kaiserslautern mehrere Personen im Datensatz haben, die verschiedensprachige Suchergebnisse bekommen, nämlich englische, russische und deutsche. Dies erlaubt es uns, den Einfluss der Regionalisierung, der durch die IP-Adresse erzeugt worden sein könnte, zu eliminieren.

Im zweiten Schritt zeigen wir zuerst die Suchergebnisse einer Person, die von Google u.a. mit Links auf russische Webseiten versorgt wird, aber mit einer deutschen IP-Adresse aus Kaiserslautern sucht.

7.2 Beispiel 2: Suche aus Deutschland mit partiell russischen Suchergebnissen

Bei unserer manuellen Inspektion fanden wir eine Person mit deutscher IP in Kaiserslautern, die für Suchbegriffe von Personen und Parteien oftmals mit einem russischen Wikipedia-Eintrag versorgt wurde. Unser Plug-In hat neben der IP-Adresse auch die jeweilig eingestellte Tastatursprache mitgespeichert – diese wurde bei dieser Person als russisch angegeben. Eine typische Suchergebnisliste für den Suchbegriff „Angela Merkel“ ist in Tabelle 12 gezeigt.

Der zweite Link verweist auf den russischsprachigen Wikipedia-Artikel von Angela Merkel, der dritte auf den englisch-sprachigen. Der vierte Link verweist auf die inhaltlich deutschsprachige Facebook-Seite von Angela Merkel. Wird der Link von jemanden geöffnet, der in Facebook eingeloggt ist und dort russisch als bevorzugte Sprache angibt, dann werden die von Facebook gestellten Textteile der Seite (z.B. die Angabe der Nutzer, etc.) auf Russisch angegeben. Dies ändert also die gezeigten Inhalte der Politikerin nicht. Zuletzt fällt noch ein zweiter englischsprachiger Link von forbes.com auf.

Diesem Nutzer werden also keine Webseiten aus Russland angezeigt, aber trotzdem scheint Google die russischsprachigen Einträge der Wikipedia für sinnvoll zu halten. Es ist unklar, warum dies der Fall ist. Wir haben bei vielen anderen Nutzern festgestellt, dass eine englische Tastatur nicht notwendigerweise mit englischsprachigen Ergebnissen verknüpft ist, daher halten wir die russische Tastaturbelegung nicht für den (einzigen) Grund dieser Suchergebnisse. Es fällt auf, dass die Person immer in ihren Google-Account eingeloggt ist – dies könnte die Quelle der Information sein.

Wir haben dann für einen Suchzeitpunkt verglichen, ob diese Person aus Kaiserslautern – die Google partiell mit russischsprachigen Webseiten versorgt – eher dieselben Links bekommt wie andere Personen aus Kaiserslautern oder wie eine andere Person mit einer ausländischen IP-Adresse, nämlich die oben schon vorgestellte Person aus Frankreich (Abbildung 30).

[18] https://twitter.com/cem_oezdemir
[19] https://www.transfermarkt.fr/cem-ozdemir/profil/spieler/170442

Tabelle 12: Typische Suchergebnisliste zum Suchbegriff "Angela Merkel" für einen Nutzer, der über den gesamten untersuchten Suchzeitraum um die 40 Suchergebnislisten eingeschickt hat und hierbei anteilig russische Ergebnisse erhalten hat.

https://de.wikipedia.org/wiki/Angela_Merkel
https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%80%D0%BA%D0%B5%D0%BB%D1%8C,_%D0%90%D0%BD%D0%B3%D0%B5%D0%BB%D0%B0
https://en.wikipedia.org/wiki/Angela_Merkel
https://ru-ru.facebook.com/angelamerkel
https://www.angela-merkel.de/
http://www.spiegel.de/thema/angela_merkel/
https://www.bundeskanzlerin.de/
https://www.forbes.com/profile/angela-merkel/
http://www.focus.de/personen/angela-merkel/
http://www.stern.de/politik/deutschland/themen/angela-merkel-4540550.html

Es ist auffällig, dass die beiden von Google mit nicht-deutschen Webseiten versorgten Personen keine Schlagzeilen bekommen. Die Person aus Kaiserslautern mit dem Verweis auf die russische Wikipedia Seite von Sahra Wagenknecht sieht auch einen Link auf ihren Twitter-Account. Der letzte Teil des entsprechenden Links "lang=ru" bewirkt nur, dass die Twitter-Seite selbst russische Wörter enthält (darunter die Übersetzung der Wörter, die zur Navigation auf Twitter genutzt werden, z.B.: "Mitteilungen", "Startseite") – die eigentlichen Tweets sind auf Deutsch. Dazu erhält die Person noch eine Überblicksseite vom Tagesspiel und einen Bunte-Artikel über Sahra Wagenknecht. Die Person in Frankreich erhält einen französischen Artikel über "Sahra la Rouge" und den englischen Wikipedia-Eintrag.

Interessanterweise bekommt der Kaiserslauterner nicht für alle Personen, die einen russischen Wikipedia-Eintrag haben, diesen auch angezeigt: Außer Dietmar Bartsch und Katrin Göring-Eckardt gibt es für alle Politikerinnen und Politiker aus unserer Menge an Suchbegriffen eine russische Wikipedia-Seite, die aber nicht immer angezeigt wird.

Abbildung 31 zeigt für denselben Suchzeitpunkt die Ergebnisse derselben vier Personen zum Suchbegriff „Cem Özdemir“. Der Kaiserslauterner bekommt hier den deutschen und den englischsprachigen Wikipedia-Eintrag, der Franzose den deutschen und den französischen Wikipedia-Eintrag. Neben den Links, die sie auch mit den konstant deutschsprachig versorgten Personen gemeinsam haben, teilen sie sich auch einen Link auf den Twitter-Account einer anderen Person mit dem Namen "Cem Özdemir", der nicht der Grünenpolitiker ist. Selbiges gilt für den jeweils letzten Eintrag, einen Link auf eine Unterseite der Top-Level-Domain „www.gruene.de“, der ansonsten weniger oft angezeigt wird. Zusammenfassend teilen sich die beiden von Google mit nicht-deutschsprachigen Webseiten versorgten Personen 6 Links. Der französische Nutzer erhält des Weiteren zwei französischsprachige Webseiten, zusätzlich einen Hinweis auf einen Fußballer und den korrekten Twitter-Account von Cem Özdemir. Die in Kaiserslautern ansässige, aber von Google partiell mit russischen Webseiten versorgte Person, erhält noch einen Nachrichtenartikel der Zeit, den die anderen nicht bekommen. Mit den anderen in Kaiserslautern ansässigen Nutzern teilt er sich 7 bzw. 6 Links.

	KL						Paris	
	DE1		DE2		RU		FR	
Zeit-angabe	Suchergebnisse	Zeit-angabe	Suchergebnisse	Zeit-angabe	Suchergebnisse	Zeit-angabe	Suchergebnisse	
Top Stories								
vor 1 Tag	https://www.zdf.de/politik/berlin-direkt/berlin-direkt---sommerinterview-vom-20-august-2017-100.html	vor 1 Tag	https://www.zdf.de/politik/berlin-direkt/berlin-direkt---sommerinterview-vom-20-august-2017-100.html					
vor 1 Tag	http://www.heute.de/das-zdf-sommerinterview-mit-sahra-wagenknecht-47797052.html	vor 1 Tag	http://www.heute.de/das-zdf-sommerinterview-mit-sahra-wagenknecht-47797052.html					
vor 23 Stund	https://www.lokalkompass.de/bochum/politik/sahra-wagenknecht-vor-gericht-d784690.html	vor 23 Stund	https://www.lokalkompass.de/bochum/politik/sahra-wagenknecht-vor-gericht-d784690.html					
Organische Suchergebnisse								
https://www.sahra-wagenknecht.de/		https://www.sahra-wagenknecht.de/		https://ru.wikipedia.org/wiki/%D0%92%D0%B0%D0%B3%D0%B5%D0%BD%D0%BA%D0%BD%D0%B5%D1%85%D1%82,_%D0%A1%D0%B0%D1%80%D0%B0		https://fr.wikipedia.org/wiki/Sahra_Wagenknecht		
https://de.wikipedia.org/wiki/Sahra_Wagenknecht		https://de.wikipedia.org/wiki/Sahra_Wagenknecht		https://de.wikipedia.org/wiki/Sahra_Wagenknecht		https://de.wikipedia.org/wiki/Sahra_Wagenknecht		
https://de-de.facebook.com/sahra.wagenknecht/		https://de-de.facebook.com/sahra.wagenknecht/		https://www.sahra-wagenknecht.de/		https://en.wikipedia.org/wiki/Sahra_Wagenknecht		
http://www.spiegel.de/thema/sahra_wagenknecht/		http://www.spiegel.de/thema/sahra_wagenknecht/		https://www.facebook.com/sahra.wagenknecht/		https://www.sahra-wagenknecht.de/		
https://www.linksfraktion.de/fraktion/abgeordnete/profil/dr-sahra-wagenknecht/		https://www.linksfraktion.de/fraktion/abgeordnete/profil/dr-sahra-wagenknecht/		http://www.spiegel.de/thema/sahra_wagenknecht/		https://www.facebook.com/sahra.wagenknecht/		
https://www.welt.de/politik/deutschland/article165927206/Diese-SPD-braucht-kein-Mensch.html		http://www.t-online.de/nachrichten/deutschland/bundestagswahl/id_81516044/sahra-wagenknecht-verteidigt-im-interview-kritik-an-asylpolitik-.html		https://www.bundestag.de/abgeordnete18/biografien/W/wagenknecht_sahra/259154		http://www.spiegel.de/thema/sahra_wagenknecht/		
https://www.welt.de/politik/deutschland/article166513610/Solche-Menschen-haben-nichts-in-der-Linken-zu-suchen.html		https://www.welt.de/politik/deutschland/article165927206/Diese-SPD-braucht-kein-Mensch.html		https://www.linksfraktion.de/fraktion/abgeordnete/profil/dr-sahra-wagenknecht/		http://tempsreel.nouvelobs.com/economie/la-crise-grecque/20150717.OBS2779/allemagne-sahra-la-rouge-l-icone-anti-merkel.html		
https://www.bundestag.de/abgeordnete18/biografien/W/wagenknecht_sahra/259154		https://www.welt.de/politik/deutschland/article166513610/Solche-Menschen-haben-nichts-in-der-Linken-zu-suchen.html		https://twitter.com/swagenknecht?lang=ru		https://www.bundestag.de/abgeordnete18/biografien/W/wagenknecht_sahra/259154		
https://www.zdf.de/politik/berlin-direkt/berlin-direkt---sommerinterview-vom-20-august-2017-100.html		https://www.bundestag.de/abgeordnete18/biografien/W/wagenknecht_sahra/259154		http://www.bunte.de/panorama/politik/sahra-wagenknecht-gestaendnis-sie-haette-so-gerne-kinder-gehabt.html		http://www.t-online.de/nachrichten/deutschland/bundestagswahl/id_81516044/sahra-wagenknecht-verteidigt-im-interview-kritik-an-asylpolitik-.html		
				http://www.tagesspiegel.de/themen/sahra-wagenknecht/		https://www.welt.de/politik/deutschland/article165927206/Diese-SPD-braucht-kein-Mensch.html		

Abbildung 30: Suchergebnislisten zum Suchbegriff "Sahra Wagenknecht" am 22.8.2017, von Nutzern mit IP-Adressen in Paris (linke Spalte) und Kaiserslautern. Gleiche Ergebnisse sind mit der gleichen Farbe dargestellt. Die ersten zwei Spalten zeigen Nutzer an, deren Suchergebnisse konstant deutsch sind, die dritte Spalte entspricht den Suchergebnissen einer Person, die über alle Suchzeitpunkte hinweg für einige der Suchbegriffe russischsprachige Webseiten angezeigt bekommt. Die vierte Spalte kommt von einem Nutzer oder einer Nutzerin mit einem IP-Standort in Paris, Frankreich.

	KL						Paris	
	DE1		DE2		RU		FR	
Zeit-angabe	Suchergebnisse	Zeit-angabe	Suchergebnisse	Zeit-angabe	Suchergebnisse	Zeit-angabe	Suchergebnisse	
Top Stories								
vor 3 Stunden	http://www.n-tv.de/politik/politik_person_der_woche/Jamaika-Aussenminister-oder-Ober-Loser-article19994815.html	vor 3 Stunden	http://www.n-tv.de/politik/politik_person_der_woche/Jamaika-Aussenminister-oder-Ober-Loser-article19994815.html					
vor 1 Tag	http://www.spiegel.de/politik/deutschland/horst-seehofer-martin-schulz-cem-oezdemir-und-katrin-goering-eckardt-kritisieren-obergrenzen-aeusserung-a-1163717.html	vor 1 Tag	http://www.spiegel.de/politik/deutschland/horst-seehofer-martin-schulz-cem-oezdemir-und-katrin-goering-eckardt-kritisieren-obergrenzen-aeusserung-a-1163717.html					
vor 9 Stunden	https://www.derwesten.de/staedte/bochum/warum-cem-oezdemir-fuer-seinen-wahlkampf-in-bochum-die-joggingpeitsche-auspackt-id211657271.html	vor 9 Stunden	https://www.derwesten.de/staedte/bochum/warum-cem-oezdemir-fuer-seinen-wahlkampf-in-bochum-die-joggingpeitsche-auspackt-id211657271.html					
Organische Suchergebnisse								
https://www.oezdemir.de/		https://www.oezdemir.de/		https://de.wikipedia.org/wiki/Cem_%C3%96zdemir		https://fr.wikipedia.org/wiki/Cem_%C3%96zdemir		
https://de.wikipedia.org/wiki/Cem_%C3%96zdemir		https://de.wikipedia.org/wiki/Cem_%C3%96zdemir		https://en.wikipedia.org/wiki/Cem_%C3%96zdemir		https://de.wikipedia.org/wiki/Cem_%C3%96zdemir		
https://de-de.facebook.com/Cem/		https://twitter.com/cem_oezdemir?lang=de		https://www.oezdemir.de/		https://www.oezdemir.de/		
https://twitter.com/cem_oezdemir?lang=de		https://de-de.facebook.com/Cem/		https://twitter.com/esekherif_		https://www.facebook.com/Cem/		
https://www.bundestag.de/abgeordnete18/biografien/O/oezdemir_cem/258834		https://www.bundestag.de/abgeordnete18/biografien/O/oezdemir_cem/258834		https://www.facebook.com/Cem/		https://twitter.com/cem_oezdemir		
https://web.de/magazine/schlagwort/cem-oezdemir		https://web.de/magazine/schlagwort/cem-oezdemir		https://www.bundestag.de/abgeordnete18/biografien/O/oezdemir_cem/258834		https://twitter.com/esekherif_		
http://www.faz.net/aktuell/politik/thema/cem-oezdemir		http://www.faz.net/aktuell/politik/thema/cem-oezdemir		http://www.faz.net/aktuell/politik/thema/cem-oezdemir		https://www.bundestag.de/abgeordnete18/biografien/O/oezdemir_cem/258834		
http://www.spiegel.de/thema/cem_oezdemir/		http://www.spiegel.de/thema/cem_oezdemir/		http://www.zeit.de/wissen/gesundheit/2017-06/cannabis-cem-oezdemir-drogen-cannabiskontrollgesetz		https://www.transfermarkt.fr/cem-ozdemir/profil/spieler/170442		
http://www.n-tv.de/politik/politik_person_der_woche/Jamaika-Aussenminister-oder-Ober-Loser-article19994815.html		http://www.n-tv.de/politik/politik_person_der_woche/Jamaika-Aussenminister-oder-Ober-Loser-article19994815.html		http://www.spiegel.de/thema/cem_oezdemir/		http://www.liberation.fr/portrait/1999/05/06/cem-ozdemir-33-ans-depute-vert-allemand-d-origine-turque-il-est-l-un-des-initiateurs-de-la-reforme-d_272246		
				https://www.gruene.de/ueber-uns/2017/cem-oezdemir.html		https://www.gruene.de/ueber-uns/2017/cem-oezdemir.html		

Abbildung 31: Suchergebnislisten zum Suchbegriff „Cem Özdemir“ am 22.8.2017, von Nutzern mit IP-Adressen in Paris (linke Spalte) und Kaiserslautern. Gleiche Ergebnisse sind mit der gleichen Farbe dargestellt. Die ersten zwei Spalten zeigen Nutzer an, deren Suchergebnisse konstant deutsch sind, die dritte Spalte entspricht den Suchergebnissen einer Person, die über alle Suchzeitpunkte hinweg für einige der Suchbegriffe russischsprachige Webseiten angezeigt bekommt. Die vierte Spalte kommt von einem Nutzer oder Nutzerin mit einem IP-Standort in Paris, Frankreich.

Sehr viel häufiger als eine Person, die russische Wikipedia-Seiten angeboten bekommt, sind Nutzer mit IP-Adressen im Inland oder Ausland, die von Google mit Mischungen aus englischen und deutschen Webseiten versorgt werden, darunter ist auch eine IP in Kaiserslautern. Wir haben uns daher für einen Suchzeitpunkt und alle Parteien die Suchergebnislisten für vier Nutzer mit IP-Adresse in Kaiserslautern angesehen. Darunter sind zwei, die nur deutschsprachige Ergebnisse bekommen (DE1 und DE2), die Person mit partiell russischen Webseiten (RU) und eine Person mit partiell englischsprachigen Webseiten, die aus den USA stammen (USA). Für alle Parteien als Suchbegriffe zeigt Abbildung 32 jeweils die Anzahl der jeweils gemeinsamen Links.

USA	12			
RU	6	10		
DE1	4	4	12	
DE2	4	4	12	12
	USA	RU	DE1	DE2

AfD

USA	10			
RU	10	10		
DE1	8	8	12	
DE2	7	7	11	12
	USA	RU	DE1	DE2

Bündnis90/Die Grünen

USA	10			
RU	6	10		
DE1	5	5	12	
DE2	5	5	12	12
	USA	RU	DE1	DE2

CDU

USA	10			
RU	4	10		
DE1	2	4	12	
DE2	2	4	12	12
	USA	RU	DE1	DE2

CSU

USA	10			
RU	7	10		
DE1	4	4	11	
DE2	4	4	12	11
	USA	RU	DE1	DE2

Die Linke

USA	10			
RU	6	10		
DE1	4	4	11	
DE2	4	4	10	11
	USA	RU	DE1	DE2

FDP

USA	12			
RU	3	10		
DE1	6	4	12	
DE2	6	4	12	12
	USA	RU	DE1	DE2

SPD

Abbildung 32: Für den 5.9.2017 und alle Parteien als Suchbegriffe wurden die Ergebnislisten von vier Nutzern mit einer IP-Adresse in Kaiserslautern miteinander auf die Anzahl gemeinsamer Links überprüft. Zwei von den Nutzern bekamen nur deutschsprachige Webseiten (DE1 und DE2), einer partiell russischsprachige und englische Webseiten (RU), einer partiell amerikanische Webseiten (USA). Gezeigt sind paarweise die Anzahl gemeinsamer Links – die Diagonale (in grau) gibt an, wie viele Links die Suchergebnisseite der entsprechenden Person enthielt.

Es fällt auf, dass die beiden Nutzer mit rein deutschsprachigen Ergebnissen fast immer identische Suchergebnislisten haben – eine Betrachtung der Listen zeigt, dass selbst die Reihenfolge fast identisch ist. Das überraschende Ergebnis ist, dass aber die beiden Personen – mit demselben IP-Standort wie die beiden anderen Nutzer–, die jeweils auch Webseiten in anderen Sprachen angezeigt bekommen, bei fünf der Parteien mehr Links gemeinsam haben als mit den anderen beiden Nutzern und bei zwei Parteien gleich viele Links miteinander gemeinsam haben wie mit den anderen beiden Nutzern.

Um dieses Muster noch zu erhärten, haben wir uns in einem letzten Vergleich für einen Suchzeitpunkt und einen Suchbegriff (22.9.2017, „Angela Merkel“) alle Suchergebnislisten von Personen mit IP-Standort im Ausland angesehen und alle Personen mit einem IP-Standort in Kaiserslautern. Da bei allen außer einer Person auch Schlagzeilen enthalten waren, konnten wir jeweils feststellen, ob Google die Person mit englischen oder deutschen Zeitangaben ("1 hour ago" vs. „Vor 1 Stunde“) ansprach; andere Sprachen in der Zeitangabe gab es in dieser Stichprobe nicht. Das erstaunliche Ergebnis zeigt Abbildung 33: Die Ähnlichkeit der Suchergebnisse ist im Wesentlichen davon abhängig, ob Google einen vermeintlich deutschsprachigen oder einen vermeintlich englischsprachigen Menschen anspricht.

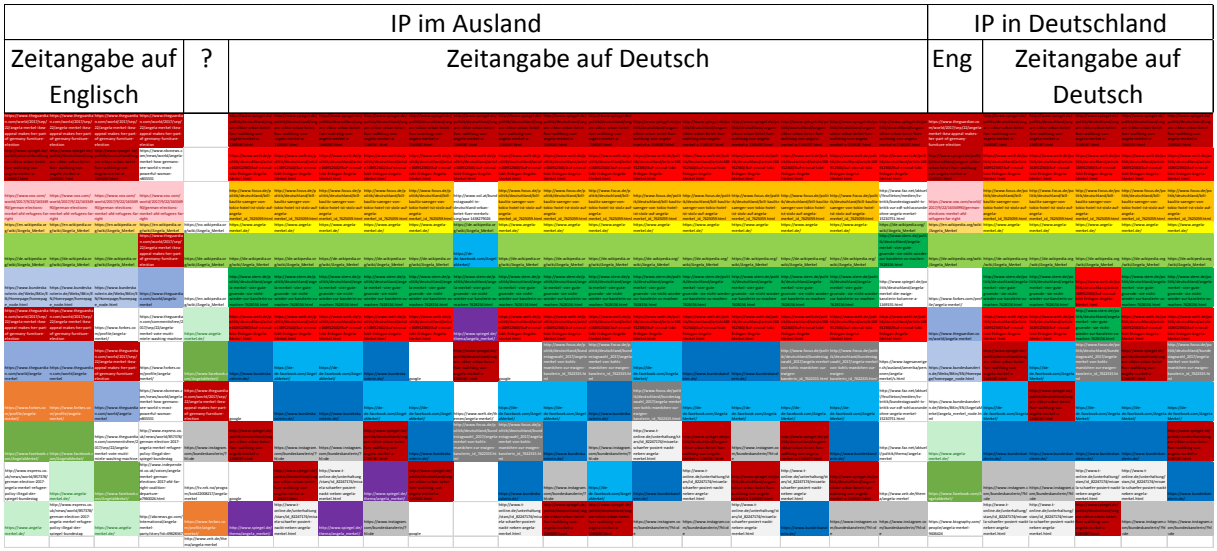


Abbildung 33: Suchergebnislisten vom 22.9.2017 um 16h zum Suchbegriff "Angela Merkel". Gezeigt sind alle Suchergebnislisten von Personen mit IP im Ausland (6x Schweiz, 3x Österreich, 2x Belgien, 2x USA, 2x Frankreich, 2x Großbritannien, Dänemark, Norwegen), jeweils in den Spalten. Zum Vergleich enthalten die letzten 6 Spalten die Suchergebnisse aller Personen mit einer IP aus Kaiserslautern. Wenn Schlagzeilen auf der Suchergebnisliste sind, gibt die Sprache der Zeitangabe ("vor 4 Stunden", "1 hour ago") Auskunft darüber, in welcher Sprache Google dem Nutzer die Webseite aufbaut. Die abweichenden Cluster werden von Zeitangaben auf Englisch begleitet, selbst bei der Person aus Kaiserslautern. Diese sind wiederum untereinander sehr ähnlich.

Obwohl die abweichenden Cluster in ihren Mustern weiterhin sehr komplex bleiben und aufgrund der eher geringen Fallzahlen von Personen mit nicht-englischen Webseiten, ist eine Trennung zwischen „deutschen“ Nutzern und „nicht-deutschen“ Nutzern deutlich zu erkennen. Es ist unklar, wie/wodurch Google entscheidet, ob eine Person deutschsprachig ist und sie nur mit deutschen Ergebnissen versorgt. Offenkundig ist aber, dass insbesondere die partiell mit englischsprachigen Ergebnissen versorgten Personen nahezu unabhängig vom Standort (in diesem Fall Österreich, Belgien, Großbritannien und Dänemark) insgesamt mit sehr ähnlichen Ergebnissen beliefert werden. Eine Ausnahme stellt eine Person aus Norwegen dar, die zu diesem Zeitpunkt keine Schlagzeilen erhielt und von der wir daher nicht wissen, in welcher Sprache Google sie anspricht. Es fällt aber auf, dass diese Person zwei norwegische Webseiten auf ihrer Suchergebnisliste hat und für einen anderen Suchbegriff zu demselben Suchzeitpunkt („Alexander Gauland“) drei Schlagzeilen mit norwegischen Zeitangaben erhielt. Eine weitere Ausnahme ist eine der Schweizer Personen, die relativ viele Nachrichten bekommt, welche die anderen nicht bekommen.

Es ergibt sich also tatsächlich ein zwar nicht isolierter, aber doch deutlich anderer Nachrichtenraum für die Personen, die zwar in Deutschland leben bzw. über einen in Deutschland befindlichen IP-Router suchen, die Google aber auf Englisch (oder in weiteren Sprachen) anspricht.

7.3 Erklärungsansatz

Am Ende ist der einfachste Erklärungsansatz vermutlich auch der richtige: Die Art der vorgefundenen, abweichenden Suchergebnislisten für deutsche IP-Standorte lässt sich schnell selbst herstellen, wenn man in seinem eigenen Google-Konto die Suchspracheinstellungen ändert. Dazu geht man auf sein eigenes Konto, steuert die Spracheinstellungen an, und stellt diese auf eine beliebige der hier genannten Sprachen, z.B. Französisch. Ein Aufruf der Google-Seite zeigt sofort, dass alle Bedienelemente nun in dieser Sprache beschriftet sind. Werden bei Suchen Schlagzeilen mit angezeigt, ändert sich die jeweilige Zeitangabe passend zur ausgewählten Sprache. Zudem zeigt sich für die Suchbegriffe dieses Projekts genau die Mischung an Suchergebnissen, die wir auch bei den abweichenden Suchergebnislisten gefunden haben: Verweise auf englische Wikipedia-Einträge, Verwechslungen von CSU, CDU, AfD und SPD mit den in den gewählten Sprachen bekannten Akronymen (z. B. „Colorada State University“, „Charles Darwin University“, „Agence Française de Développement“, „symphysis pubis dysfunction“), und deutschen Webseiten, die tatsächlich die entsprechenden deutschen Parteien beschreiben oder ihnen gehören oder ihre Social-Media-Accounts. Die Übereinstimmung mit den gefundenen Abweichungen ist also groß – trotzdem kann dies nicht beweisen, dass diese Sprachfilterung der einzige Mechanismus ist, wie es zu abweichenden Suchergebnislisten kommt. Wir folgen hier aber „Ockham‘s razor“, der Forderung, dass ein Resultat immer mit dem einfachsten verfügbaren Modell erklärt werden soll, dass es beschreibt. Wir schlussfolgern daher, dass die abweichenden Resultatlisten im Wesentlichen durch die von den Datenspendern und -spenderinnen selbst eingestellte, im Profil hinterlegte Suchsprache begründet werden. Das Ergebnis ist eine Mischung aus Sprachfilter, Regionalisierung (vermutlich durch den IP-Standort) und den relevantesten Webseiten zum Suchbegriff.

Es ist weiterhin wichtig zu bemerken, dass die hier gezeigten Fälle an keiner Stelle flächendeckend Nachrichten enthielten, die eine andere politische Perspektive auf die Personen oder Parteien darstellten. Die von Eli Pariser gefürchtete **inhaltliche** Filterblase, in der die einen eine ganz andere politische Perspektive auf Personen oder Inhalte bekommen als andere, konnten wir in den manuell inspizierten Fällen **nicht** entdecken. Englischsprachige Nachrichten kamen z. B. von The Guardian, französischsprachige von Le Figaro, die russischsprachigen waren Wikipedia bzw. Wikimedia-Seiten. In vielen Fällen unterschieden sich die Suchergebnislisten der als „ausländisch“ erkannten Nutzer in der genauen Quelle von aggregierten Themenseiten, wie z. B. http://www.faz.net/aktuell/politik/thema/martin-schulz von den Suchergebnissen der stärker regionalisierten oder mit Nachrichten versorgten und als „deutschsprachig“ erkannten Nutzern.

Trotzdem stellt es angesichts der vielen Minderheiten eines Landes, die in erster oder zweiter Generation aus anderen Ländern stammen, eine wichtige, generelle Frage dar, welche Informationen diese von wo erhalten sollten, damit sich die Diskussionsräume weitestgehend überlappen.

8 Zusammenfassung

Für die Datenspenderinnen und Datenspender können wir die klare Schlussfolgerung ziehen, dass sich die allermeisten Suchergebnislisten für alle Suchbegriffe hinreichend überlappen und wir alleine deswegen – unabhängig von den Inhalten – die algorithmenbasierte Bildung und Verhärtung von isolierenden Filterblasen verneinen können. Dabei ist festzuhalten, dass Personen, die Google auch mit fremdsprachigen Webseiten versorgt (vermutlich durch eine andere Suchspracheinstellung dieser Nutzer ausgelöst), in vielen Fällen immer noch eine gewisse Überlappung der Suchergebnisse erfahren, trotzdem könnte man hier von einer Filterblase sprechen. Auch wenn diese sprachbedingt unvermeidbar ist, bedeutet dies, dass man basierend auf den vier Mechanismen, die zusammenkommen müssen, um gefährliche Filterblasen zu erzeugen, in die Inhalte sehen muss. Manuelle Inspektionen und auch eine Vielzahl der ausgelieferten Top-Level-Domains zeigen eher Mainstream-Inhalte auf, es gibt aber auch vereinzelte Top-Level-Domains und URLs, die weniger leicht einzuordnen sind. Hier bedarf es der interdisziplinären Analyse der vorliegenden Daten und möglichst auch weiterer Studien.

Wie steht es mit der Frage, ob der Algorithmus uns alle mit Inhalten versorgt, die einseitig sind? Zuerst einmal spricht die große Vielfalt der insgesamt ausgelieferten Top-Level-Domains dagegen, aber das beantwortet weder die Frage, ob es nicht noch andere Top-Level-Domains gegeben hätte, die mit angezeigt hätten werden sollen, noch die Frage, ob manche von diesen Top-Level-Domains übermäßig häufig oder selten angezeigt werden. Womit sollte man dies vergleichen? Dazu könnte man – als eine mögliche Dimension – die Beliebtheit der Top-Level-Domain unter allen Internetnutzern als Grundlage nutzen, wie sie beispielsweise im Nielsen-Online-Panel fortlaufend gemessen wird. Das Nielsen-Panel beruht auf einer repräsentativen Nutzergruppe, deren Surfverhalten mit einer Software minutiös festgehalten wird. Für jede Top-Level-Domain ist auf Basis der Panelmessung der sogenannte „active reach“ bekannt, der prozentuale Anteil von Personen, die mindestens einmal in einem definierten Zeitraum auf dieser Top-Level-Domain waren[20].

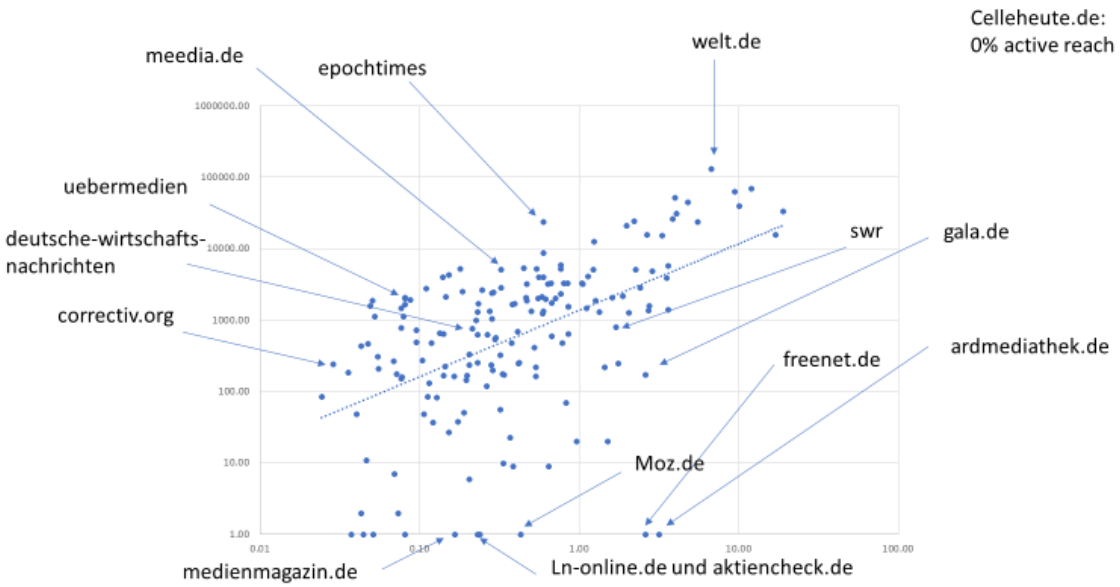


Abbildung 34: Das Diagramm zeigt für verschiedene Top-Level-Domains jeweils den im Nielsen-Report vom August festgestellten "active reach" (Prozent der Internetnutzer, die mindestens einmal eine URL der Top-Level-Domain in diesem Zeitraum besucht hatten) und die Anzahl der Suchergebnisse im August 2017 von dieser Top-Level-Domain. Es handelt sich um eine doppelt-logarithmische Auftragung. Es wurden einige Quellen mit starker Abweichung von der Trendlinie gekennzeichnet.

[20] Bei sehr kleinen Fallzahlen sind die Werte allerdings nicht mehr aussagekräftig, wenn z.B. weniger als eine Handvoll Nutzer diese Webseiten besuchten. Dies wird im Nielsen Report jeweils kenntlich gemacht.

Hier handelt es sich allerdings wieder um ein Henne-Ei-Problem, da die Marktführerschaft unter den Suchmaschinen in Deutschland eindeutig bei Google liegt und eine Vielzahl von Webbesuchen über Google vermittelt werden. Das heißt, in den Zahlen liegt ein unbekannter Anteil von Besuchen, die überhaupt erst über Google erreicht werden, und damit stellen die Zahlen kein von Googles Suchmaschinenalgorithmus unabhängiges Beliebtheitsmaß dar. Daher sind die im Folgenden festgestellten Befunde eher Hinweise auf weitere Forschungsfragen als Resultate des Datenspendeprojektes.

Abbildung 34 zeigt die Auftragung des "active reach" (der Reichweite) einer Domain gegen die im selben Zeitraum von Google an die Datenspenderinnen und -spender gelieferten URLs dieser Domain aus den Schlagzeilen (nicht den organischen Suchen). Da die Unterschiede pro Achse sehr groß sind (von 0.01 bis 100 % bzw. von 1 bis 1.000.000 angezeigte URLs), wurde eine doppelt logarithmische Auftragung gewählt. Das Muster der Daten erlaubt es, durch diese Auftragung eine lineare Regression zu legen. Die Formel $y = 1.373 * \text{active reach}^{0.9}$ besagt im Wesentlichen, dass die Anzahl der gelieferten URLs fast linear mit der Reichweite korreliert, mit einem kleinen Dämpfer für größere Reichweiten.

In der Abbildung werden einige Punkte herausgegriffen, die von diesem Muster mit ungefähr Faktor 5 nach oben oder unten abweichen. D.h., sie werden mindestens 5-mal mehr oder 5-mal weniger als nach der Formel erwartet ausgeliefert. Dazu gehören auf der einen Seite Medien, die über Medien berichten, wie correctiv.org, meedia.de und uebermedien. Correctiv.org ist eine Korrekturwebseite, die Fehlmeldungen aufgreift und richtigstellt. Während diese Top-Level-Domain im NielsenPanel im August keine Reichweite hat, wird eine seiner URLs an 245 Personen im Datenspendeprojekt ausgeliefert, fast fünfmal mehr als erwartet. Trotzdem ist das insofern weniger bemerkenswert als zuerst gedacht, da es um genau eine Schlagzeile geht, die zu diesem Suchzeitpunkt (24.8.2017, 12h, Suchbegriff „Alexander Gauland“) an alle Spender geliefert wird. Da wir schon festgestellt haben, dass die Suchergebnisseiten insgesamt sehr einheitlich sind, betrifft eine solche Entscheidung also immer alle zum Zeitpunkt Suchenden. Es handelt sich also um ein einmaliges Ereignis, das schlagartig die Anzahl der ausgelieferten URLs über den erwarteten Wert hebt. Artikel von meedia und uebermedien werden 10 bzw. 16-mal mehr ausgeliefert als erwartet. Es ist nochmals wichtig zu betonen, dass der Vergleich mit den Nielsen-Reichweiten methodisch schwierig ist: Weder stellt der "active reach" ein von Googles Suchmaschine unabhängiges Beliebtheitsmaß dar noch kann die Beliebtheit uneingeschränkt als direktes Maß für die inhaltliche Qualität der Beiträge der Top-Level-Domains gelten.

Trotzdem mag es eine interessante Anfangsbeobachtung sein, dass die 10 am stärksten überrepräsentierten Quellen (bzgl. des Erwartungswertes durch die Regression) die folgenden sind:

1. Epochtimes (mit insgesamt 240.004 Links in den Suchergebnissen, erwartet 840)
2. Chiemgau24.de (1.905 Links, erwartet 22)
3. Jungewelt.de (1.595 Links, erwartet 81)
4. Stimme.de (5.259 Links, erwartet 274)
5. Ka-news.de (4.047 Links, erwartet 218)
6. Tichyseinblick.de (4302 Links, erwartet 235)
7. Welt.de (131.793 Links, erwartet 8120)
8. Übermedien.de (2102 Links, erwartet 131)
9. Cicero.de (2785 Links, erwartet 174)
10. Butenunbinnen.de (2058 Links, erwartet 131)

Wie immer bei einer relativen Bewertung sind die meisten der „Ausreißer“ Quellen, die insgesamt nicht oft in den Suchergebnissen auftauchen. Im Wesentlichen sind aber die reichweitenstärksten Medien auch in den von Google ausgelieferten Schlagzeilen vertreten. Zusammen mit den in den organischen Suchergebnissen angezeigten Webseiten der Parteien und Personen, den Wikipedia-Einträgen und den Nachrichten zeigt sich ein breiter Blick auf die Parteien und Politiker, der insgesamt eine umfassende Information und Meinungsbildung erlaubt. Zusätzlich sollte erwähnt werden, dass natürlich weitere

Suchbegriffkombinationen mit den hier untersuchten Suchbegriffen jederzeit eine detailliertere Suche erlauben würden, falls die erste Suchergebnisliste unbefriedigend wäre.

8.1 Generalisierung der Ergebnisse: Möglichkeiten und Grenzen

Die gezeigten Ergebnisse können nicht uneingeschränkt auf alle Google-Nutzer in Deutschland übertragen werden, da die Datenspenderinnen und -spender kein repräsentatives Sample darstellen (bedingt durch die eigene Entscheidung zur Teilnahme). Interessanterweise könnten aber gerade diejenigen Spender, die anscheinend eine andere Suchsprache eingestellt hatten, das Argument entkräften, dass die Nutzermenge so homogen war, dass alleine dadurch der geringe, beobachtete Personalisierungsgrad erklärbar sei. Es erscheint unwahrscheinlich, dass alle englischsprachigen Nutzer aus den verschiedensten Ländern homogen hinsichtlich ihrer Personalisierung sind und trotzdem zeigt Abbildung 33 eine bemerkenswerte Homogenität ihrer Suchergebnislisten. Auch hier folgen wir Ockham‘s razor und nehmen an, dass die Suchergebnislisten im Wesentlichen bestimmt sind von der Suchspracheneinstellung, dem Standort (z. B. durch IP-Adresse) und der allgemeinen Relevanz der Webseiten, in geringem Maß gefolgt von schon einmal durch das Individuum besuchten Webseiten und vielleicht auch personalisierten Links. Insofern würde es uns überraschen, wenn weitere Studien stark abweichende Personalisierungsgrade zeigen würden.

In jedem Fall ist die Beobachtung nur gültig für den begutachteten Zeitraum und die untersuchten Suchbegriffe. Es kann nicht ausgeschlossen werden, dass sich der Personalisierungsgrad über die Zeit ändert, aber die Studie zeigt, wie dieser einfach, kostengünstig, bezüglich der Suchbegriffe flexibel, vertrauenswürdig und automatisiert überwacht werden kann. Dies gilt für alle Suchmaschinen.

8.2 Forderung nach geeigneten Schnittstellen zur Überprüfung der Filterblasentheorie in sozialen Netzwerken und auf anderen Intermediären

Für soziale Netzwerke wie Facebook, andere soziale Medien wie Twitter, Instagram oder YouTube gilt dies allerdings nicht. Es wäre momentan nur mit größtem Aufwand möglich, beispielsweise alle politischen Anzeigen eines Nutzers im Newsfeed seines Facebook-Accounts herauszufiltern und zentral zu sammeln. ProPublica hat solch ein Projekt während der Bundestagswahl 2017 versucht. Eine API, die den selektiven Zugriff auf diese Inhalte erlaubt, gibt es nicht. Daher bestand nur die Wahl, Personen aufzufordern, vollen Zugriff zu erlauben (nicht vertrauenswürdig) oder die Personen zu bitten, Screenshots von politischen Anzeigen einzureichen. Der zweite Weg ist natürlich weder einfach, noch kann die Analyse automatisch erfolgen, da die Inhalte erst per Hand beschrieben und damit durchsuch- und zusammenfassbar gemacht werden können.

Es ergibt sich daraus die Folgerung, dass die notwendigen Schnittstellen zu allen sozialen Medien und sozialen Netzwerken geschaffen werden müssen, deren Personalisierungsgrad zur algorithmenbasierten Bildung oder Verhärtung von Filterblasen führen könnte. Eine geeignete Kontrolle des Personalisierungsgrades enthebt die entsprechenden Firmen – solange dieser niedrig bleibt – von Forderungen der Gesellschaft, weitere Einblicke in ihren Softwarecode zu erhalten und begegnet gesellschaftlichen Sorgen[21]. Sollte sich ein hoher Personalisierungsgrad (auf relevanten Inhalten) und ein gewisser Grad an Isolation der Filterblasen zeigen, bedarf es dann weiterer Untersuchungen, inwieweit die Inhalte extrem unterschiedliche Perspektiven aufweisen und den gesellschaftlichen Diskurs erschweren. Mit einem solch gestaffelten Modell der Einblickstiefe durch Black-Box-Analysen in Abhängigkeit vom gemessenen Personalisierungsgrad wäre also beiden Seiten geholfen.

[21] Dies bezieht sich natürlich nur auf die Bedenken zur Bildung und Verhärtung von Filterblasen. Weitere problematische Effekte von sozialen Medien oder sozialen Netzwerken bedürfen ihrer jeweils eigenen Kontrolle und Überwachung.

9 Danksagung

Wir bedanken uns an dieser Stelle bei den Landesmedienanstalten Bayern (BLM), Berlin-Brandenburg (mabb), Hessen (LPR Hessen), Rheinland-Pfalz (LMK), Saarland (LMS) und Sachsen (SLM), die das Projekt gefördert haben, sowie bei Spiegel Online als unserem Medienpartner. Wir möchten dabei unseren großen Dank an Uwe Conradt, Direktor der LMS, aussprechen, der das Projekt vorgeschlagen hat. Danke auch an Dr. Anja Zimmer, Direktorin der mabb und Siegfried Schneider, Präsident der BLM, die sich zusammen mit Uwe Conradt besonders für dieses Projekt eingesetzt haben und es damit erst ermöglicht haben. Unser besonderer Dank für eine großartige Projektbetreuung geht an Adrian Gerlitsch von der BLM, der immer und unermüdlich konstruktive Lösungen für alle Probleme gefunden hat.

10 Quellen

(Bozdag & van den Hoven, 2015) Bozdag, E., & van den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. Ethics and Information Technology, 17(4), 249-265.

(Horchert, 2017) Horchert, J (2017). "Googeln für die Wissenschaft", SpiegelOnline, 6.7.2017, http://www.spiegel.de/netzwelt/web/datenspende-forschungsprojekt-ueber-google-sucht-freiwillige-helfer-a-1156060.html (abgerufen am 15.5.2018)

(Krafft et al., 2017) Tobias D. Krafft, Michael Gamer, Marcel Laessing, Katharina A. Zweig. Filterblase geplatzt? Kaum Raum für Personalisierung bei Google-Suchen zur Bundestagswahl 2017. Technical Report. https://doi.org/10.13140/RG.2.2.29139.07203

(Moe & Syvertsen, 2007) Moe, H., and Syvertsen, T. (2007). "Media institutions as a research field: Three phases of Norwegian broadcasting research.", Nordicom Review, Jubilee Issue 2007, pp. 149-167.

(Pariser, 2011) Eli Pariser. The Filter Bubble: What the Internet Is Hiding from You. Penguin Press, New York, 2011, ISBN 978-1-59420-300-8.

(Schmidt et al., 2017) Jan-Hinrik Schmidt, Lisa Merten, Uwe Hasebrink, Isabelle Petrich, Amelie Rolfs. Zur Relevanz von Online-Intermediären für die Meinungsbildung. Arbeitspapiere des Hans-Bredow-Instituts Nr. 40, Hans-Bredow-Institut, Hamburg, 2017. https://www.hans-bredow-institut.de/uploads/media/default/cms/m dia/67256764e92e34539343a8c77a0215bd96b35823.pdf

(Statista, 2016) Statista (2016). Anzahl der Suchanfragen bei Google weltweit in den Jahren 2000 bis 2016 (in Milliarden). https://de.statista.com/statistik/daten/studie/71769/umfrage/anzahl-der-google-suchanfragen-pro-jahr/ (abgerufen am 15.5.2018)

(Statista, 2018) Statista (2018). Marktanteile der führenden Browserfamilien an der Internetnutzung in Deutschland von Januar 2009 bis März 2018. https://de.statista.com/statistik/daten/studie/13007/umfrage/marktanteile-der-browser-bei-der-internetnutzung-in-deutschland-seit-2009/ (abgerufen am 15.5.2018)

(Teevan et al., 2011) Teevan, J., Karlson, A., Amini, S., Brush, A. J., & Krumm, J. (2011, August). Understanding the importance of location, time, and people in mobile local search behavior. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (pp. 77-80). ACM.

(Vīķe-Freiberga et al., 2013) Vīķe-Freiberga, V., Däubler-Gmelin, H., Hammersley, B., & Maduro, L. M. P. P. (2013). A free and pluralistic media to sustain European democracy. Report, High Level Group on Media Freedom and Media Pluralism. http://ec.europa.eu/information_society/media_taskforce/doc/pluralism/hlg/hlg_final_report.pdf (abgerufen am 15.5.2018)

(Weare, 2018) Weare, C. B. (2009). "System and Method for Personalized Search", US Patent US7599916, Patent der Microsoft Corporation, Redmond, WA, USA, erteilt am 6. Oktober 2009.

(Weisberg, 2011) Weisberg, J. (2011). “Bubble Trouble- Is Web personalization turning us into solipsistic twits? Slate, 10.06.2011, http://www.slate.com/articles/news_and_politics/the_big_idea/2011/06/bubble_trouble.html (abgerufen am 12.5.2018)

(Zuiderveen Borgesius et al., 2016) Zuiderveen Borgesius, F., Trilling, D., Moeller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should We Worry About Filter Bubbles? Internet Policy Review. Journal on Internet Regulation, Volume 5, Issue 1.

(Zweig et al., 2017) Zweig, K. A., Deussen, O., & Krafft, T. D. (2017). Algorithmen und Meinungsbildung. Informatik-Spektrum, 40(4), 318-326.

Appendix

A Kategoriensystem

Das Kategoriensystem dient der Zuordnung von Domains, die im Rahmen der Datenspende angezeigt wurden. Es wurde von der Bayerischen Landeszentrale für neue Medien (BLM) entwickelt, die auch die Codierung vorgenommen hat.

Hauptkategorien

Jede Domain wird einer Hauptkategorie zugeordnet. In Konfliktfällen ist die prägnantere Zuordnung zu treffen.

a) Owned Content

Zur Kategorie Owned Content gehören alle Kanäle bei denen Parteien ihre Meinung ungefiltert darstellen kann, sei es über eine persönliche Webseite eines Parteimitgliedes eines Ortsverbandes oder der Partei selbst. Social-Media-Kanäle gehören faktisch ebenfalls dieser Kategorie an, werden jedoch für dieses Kategorienschema separat in einer eigenen Kategorie codiert.

b) Social Media

Unter Social Media sind alle Kanäle zu verstehen, die sich durch ihren sozialen Charakter auszeichnen und die Interaktion zwischen Anbieter und Nutzer als Fokus setzen. Die Domains bieten die Möglichkeit der Vernetzung; sich also untereinander auszutauschen und mediale Inhalte einzeln oder in einer definierten Gemeinschaft oder offen in der Gesellschaft zu erstellen und weiterzugeben. Beispiele für soziale Netzwerke, die unter diese Kategorie fallen sind Facebook, Twitter oder Instagram.

c) Wikipedia

Diese Kategorie wird an alle Wikipedia Domains vergeben.

d) Medien

Die Domain wird dieser Kategorie zugeordnet, wenn sie von einem Medienanbieter betrieben wird. Als Medienanbieter werden Angebote mit publizistischem Inhalt verstanden, die entweder Onlineableger klassischer Printmedien sind, von privaten Fernseh- oder Radioangeboten oder Ableger öffentlich-rechtlicher Angebote. Auch ausländische publizistische Angebote sowie Online Only Angebote werden zu dieser Kategorie gezählt (siehe Unterkategorien). Beispiele sind: www.schwarzwaelder-bote.de, news.sky.com, www.news.de.

e) Freemail Portale

Freemail Portale stellen kostenlos E-Mail-Adressen sowie E-Mail-Postfächer zum Senden und Empfangen von E-Mails bereit. Darüber hinaus finden sich auf diesen Seiten häufig Nachrichtenspalten. Die gängigen Freemail Portale, die in diese Kategorie codiert werden sind z.B. web.de, yahoo.com oder gmx.net. Freemail Angebote, die in den Ergebnissen der “Google News” erscheinen, werden nur der Hauptkategorie zugeordnet, keiner Unterkategorie.

f) Öffentlich finanziert

Zur Kategorie Staatliche Institutionen gehören alle öffentlichen Stellen die von öffentlichen Gelder finanziert werden, wie zum Beispiel die Bundeszentrale für politische Bildung (bpb) sowie die Webauftritte von Städten und Gemeinden.

g) Sonstige

Alle Domains, die nicht eindeutig einer Kategorie zugeordnet werden können, werden in die Kategorie Sonstige codiert.

Unterkategorien

Unterkategorien werden immer dann an Domains vergeben, wenn bereits die Hauptkategorie „Medien“ vergeben wurde. Alle anderen Ausprägungen der Hauptkategorie werden nicht weiter unterkategorisiert. Es wird pro Domain nur eine Unterkategorienausprägung vergeben.

a) Print

Die Kategorie umfasst alle Onlineangebote, die Onlineableger klassischer deutscher Printmedien (Tageszeitungen und Zeitschriften) sind. Es werden Domains erfasst, die sowohl deutschen Ursprungs sind als auch ihre Inhalte auf Deutsch publizieren. Beispiele sind: spiegel.de, faz.de, welt.de.

b) TV

Es werden alle Angebote codiert, die Onlineableger zugelassener deutscher privater Fernsehanbieter sind. Sowohl Ursprung als auch Sprache der publizierten Inhalte ist Deutsch. Beispiele sind: rtl.de, prosieben.de.

c) Öffentlich-Rechtliche (ÖRR)

In diese Kategorie fallen alle Onlineangebote des deutschen öffentlich-rechtlichen Rundfunks, wie z.B. deren Onlineableger der Hörfunk- und Fernsehprogramme. Beispielhaft können folgende Angebote dieser Kategorie zugeordnet werden: daserste.de, ndr.de, zdf.de.

a) Online Only

Medienangebote, deren einziger Verbreitungsweg das Internet ist, werden in diese Kategorie codiert. Beispiele hierfür sind: 02elf.net oder www.promiflash.de.