

Digital Business University of Applied Sciences

Data Science und Business Analytics

ADS 51 – ADS-05: Visualisierung und Kommunikation

Prof. Dr. Marcel Hebing

**Vergleich der Trends von Medienpräsenz und Umfragewerten  
deutscher Parteien**

Studienarbeit

Eingereicht von: Mareike Lass-Hennemann

Matrikelnummer: 190190

25.01.2025

## **Inhaltsverzeichnis**

<b>Abbildungsverzeichnis.....</b>	<b>3</b>
<b>Executive Summary .....</b>	<b>3</b>
<b>Einleitung .....</b>	<b>4</b>
<b>Daten und Methoden .....</b>	<b>4</b>
<b>Ergebnisse.....</b>	<b>6</b>
<b>Diskussion und Handlungsempfehlung .....</b>	<b>10</b>
<b>Literaturverzeichnis / Quellen .....</b>	<b>11</b>

## **Abbildungsverzeichnis**

Abbildung 1: Wie häufig kommen die Parteibezeichnungen insgesamt vor? .....	6
Abbildung 2: Anteil wordcounts vs. Umfragewerte nach Parteien.....	7
Abbildung 3: Anteil wordcounts vs. Umfragewerte CDU sowie Trends.....	7
Abbildung 4: Anteil wordcounts vs. Umfragewerte SPD sowie Trends.....	8
Abbildung 5: Anteil wordcounts vs. Umfragewerte Grüne sowie Trends.....	8
Abbildung 6: Anteil wordcounts vs. Umfragewerte FDP sowie Trends.....	9
Abbildung 7: Anteil wordcounts vs. Umfragewerte AfD sowie Trends.....	9

## **Executive Summary**

Im Projekt „Vergleich der Trends von Medienpräsenz und Umfragewerten deutscher Parteien“ soll untersucht werden, inwiefern die mediale Präsenz der Parteien, gemessen anhand des Vorkommens ihrer Bezeichnungen (wordcounts) in verschiedenen digitalen deutschen Zeitungen sich in den Umfragewerten der Sonntagsfrage (abgerufen von wahlrecht.de) spiegeln lässt. Folgen einer hohen medialen Präsenz hohe Umfragewerte? Festzustellen ist, dass die Trends der Präsenz und der Umfragewerte sich für die Parteien jeweils durchaus in ähnliche Richtungen entwickeln, aber nicht immer die Umfragewerte den wordcounts eindeutig folgen. Herauszuhoben sind die sich positiv entwickelnden Verläufe der Umfragewerte von CDU/CSU und AfD. Wobei die CDU/CSU als sowieso bekannte, große Partei scheinbar eher von einer niedrigen medialen Präsenz, die AfD als kleine, rechtspopulistische Partei dagegen von einer besonders hohen medialen Präsenz profitiert.

## **Einleitung**

Es sind bereits verschiedene Untersuchungen und Studien zum Einfluss der Medien auf Wählerverhalten vorhanden. Aus einer deutschen Untersuchung zur Wahlbeteiligung lässt sich schließen, dass Medienpräsenz zwar einen Einfluss auf die Wahrnehmung von Parteien haben kann, dieser sich aber nicht direkt in Wahlergebnissen widerspiegelt. Vielmehr spielen Faktoren wie langfristige Einstellungen der Wähler, Mobilisierungsstrategien der Parteien und die Art der Mediennutzung beider Seiten lt. Schulz (2023) eine wichtige Rolle bei der Wahlentscheidung.

Eine Untersuchung aus Spanien zur Beziehung zwischen der Sichtbarkeit von politischen Führungspersonen während Wahlkämpfen und den Wahlergebnissen, deutet dagegen darauf hin, dass eine erhöhte Medienpräsenz die Wahlergebnisse insbesondere von kleinen Parteien aus dem rechtspopulistischen Spektrum durchaus positiv beeinflussen kann (Estevez, Dominguez & Graña, 2019).

In diesem Projekt soll nun untersucht werden, inwiefern sich die Trends der medialen Präsenz und der Umfragewerte zur Sonntagsfrage in ähnliche Richtungen entwickeln. Es wird zudem angenommen, dass eine steigende/sinkende mediale Sichtbarkeit die Bekanntheit und Wahrnehmung von Parteien steigern/vermindern kann, was sich dann in den Umfragewerten widerspiegeln könnte. Dafür müsste der Trend der Umfragewerte dem Trend der Präsenz nachfolgen.

Die Rohdaten für das Projekt wurden mittels Web Scraping gewonnen. Dies ist ein automatisierter Prozess, der html- oder andere strukturierte Datenquellen aufruft und Text-Informationen sammelt. Die extrahierten Text-Daten können dann für verschiedene Zwecke wie die Analyse von Markttrends oder Forschung genutzt werden (Wikipedia, 2022).

Zur Analyse der Umfragewerte wurden gesammelte Ergebnisse der Sonntagsfrage, veröffentlicht von [wahlrecht.de](http://wahlrecht.de), herangezogen.

## **Daten und Methoden**

Freundlicherweise wurden die Rohdaten zu diesem Projekt zur Verfügung gestellt von Prof. Dr. Hebing über den DBU-Sharing-Point. Sie stammen aus dem täglichen Webscraping von 59 Webseiten und werden seit April 2021 gesammelt. Für dieses Projekt

werden nur Webseiten aus dem Inland, die politisch berichten, ausgewählt und zwar:  
Süddeutsche Zeitung, Die Zeit, Frankfurter Allgemeine, Der Tagesspiegel, Die  
Tageszeitung, Hamburger Abendblatt, Berliner Zeitung, Welt, Eßlinger Zeitung,  
Handelsblatt, ntv, The Pioneer, Südwest Presse, Stuttgarter Nachrichten,  
Deutschlandfunk, Deutsche Welle, Spiegel Online, Manager Magazin, Stern, Tagesschau,  
Wirtschaftswoche, Watson.

Von jeder Zeitung liegt nach herunterladen und entpacken der .tar-Dateien vom  
Sharepoint pro Tag zwischen dem 01.04.2021 bis zum 30.11.2024 je eine HTML-Datei vor.  
Der Textinhalt wird mithilfe des Moduls BeautifulSoup ausgelesen. Über Pandas Funktion  
.value\_counts() werden alle Wörter gezählt, sofern diese nicht in einer zuvor geladenen  
Stopp-word-Liste enthalten sind (stop-words, 2016). Auf Basis der gezählten Wörter wird  
die Tabelle „count“ in die Datenbank db.sqlite3 überschrieben. Diese enthält 41.588.493  
Fälle mit den vier Attributen count (Wie oft vorgekommen?), word (Welches Wort?),  
paper (Welches Medium?), date (Welches Datum?).

Für die Analyse wird zunächst eine SQL-Abfrage auf die Tabelle „count“ vorgenommen, die  
nach Parteibezeichnungen in Form der Wörter „cdu“, „cdu/csu“, „spd“, „grünen“, „fdp“  
und „afd“ sucht. Es werden 42.473 Fälle zur Analyse gefunden und in den Dataframe „df“  
mit der Struktur der oben beschriebenen Tabelle „count“ geschrieben. Es sind also zu  
jedem Fall wieder die Attribute „word“, „count“, „date“, „paper“ vorhanden. Aus der  
Spalte „date“ werden neue Attribute „year“, „month“ und „year\_month“ abgeleitet.  
Zusätzlich wird eine neue Spalte „partei“ eingerichtet, welche die gesuchten  
Parteibezeichnungen nach einer definierten Liste zur jeweiligen Partei sortiert. Mit diesem  
Dataframe kann ein count-plot erstellt werden, welcher darstellt, wie häufig die  
Bezeichnungen der Parteien jeweils insgesamt vorkommen. Nun werden die Zählungen  
summiert und gruppiert nach Jahr, Monat und Partei in einen weiteren Dataframe  
„monthly\_data“ gespeichert. Für den später geplanten Vergleich mit den Umfragewerten  
wird an dieser Stelle eine monatliche Gesamtsumme an wordcounts sowie der jeweilige  
prozentuale Anteil an den Gesamt-wordcounts jeder Partei berechnet. Der so optimierte  
Dataframe wird für eine Matplotlib-figure verwendet, die die Entwicklung der wordcounts  
jeder Partei im Jahresverlauf darstellt.

In einem weiteren Schritt werden die Umfragewerte der Sonntagsfrage, heruntergeladen von wahlrecht.de und gespeichert in einer Excel-Datei mit einem Arbeitsblatt pro Umfrageinstitut, in Dataframes geladen und verarbeitet. Es liegen Daten der Institute Allensbach, Verian, Forsa, Forschungsgruppe Wahlen, infratest dimap, INSA und yougov vor. Variablen, die im Dataframe df\_gesamt aller Institute mit 2041 Zeilen vorhanden sind, sind: datum (Zeitpunkt der Veröffentlichung der Umfragewerte), befragte (Anzahl), institut, cdu/csu, spd, grüne, fdp, afd (die Partei-Attribute enthalten die jeweiligen Umfragewerte in Prozent). Es wird für die Darstellung ein weiterer Dataframe df\_aggregiert erstellt, welcher für jede Partei einen Mittelwert aus den monatlich aufsummierten Werten aller Institute enthält sowie das Datum, aber nicht mehr die Attribute „befragte“ und „institut“. Auf diese Weise können lineplots der Umfragewerte für jede Partei im Jahresverlauf erstellt und mit denen der wordcounts verglichen werden. Für einen übersichtlicheren visuellen Vergleich werden zudem mithilfe der numpy-polyfit-Funktion (3.Polynom) Trendlinien für die wordcounts und die Umfragewerte modelliert. Als Trend ist hier die unabhängig von Schwankungen beobachtete Grundrichtung der Zeitreihe zu verstehen. Es wird außerdem für jede Partei eine durchschnittliche Differenz zwischen den Trends berechnet.

## Ergebnisse

In Abb. 1 wird die Anzahl der counts der gesuchten Parteibezeichnungen dargestellt. Alle Parteien bewegen sich im Bereich zwischen 7500 bis 10.000 counts.

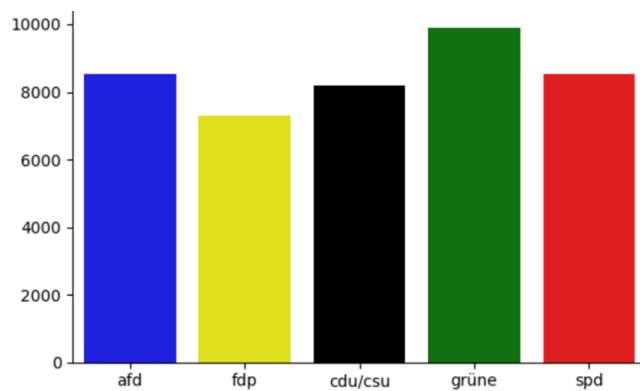


Abbildung 1: Wie häufig kommen die Parteibezeichnungen insgesamt vor?

In Abbildung 2 werden die prozentualen Anteile der wordcounts und die prozentualen Mittelwerte zur Sonntagsfrage jeweils aufgeteilt nach den Parteien im Verlauf der Jahre 2021 bis 2024 dargestellt. Es zeigt sich, dass die AfD besonders ab Mai 2023 einen deutlich größeren Anteil der wordcounts für sich einnimmt und auch bei den Umfragewerten ab

Juni 2023 an zweiter Stelle steht. Im Gegensatz dazu hat z.B. die CDU über alle Jahre einen eher niedrigen Anteil an den wordcounts, erreicht aber dennoch durchgehend im Vergleich die höchsten und auch steigende Umfragewerte.

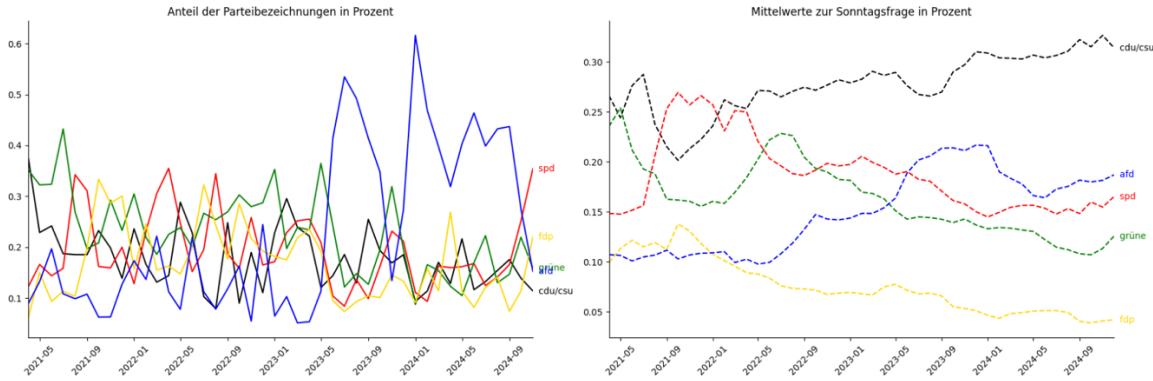


Abbildung 2: Anteil wordcounts vs. Umfragewerte nach Parteien

Die Abbildungen 3 bis 8 zeigen im Detail zu jeder Partei die Daten zur Sonntagsfrage im Vergleich zu den wordcounts sowie jeweils den Trend.

Wie schon in Abbildung 2 angedeutet, zeigen sich für die CDU gegenläufige Trends der Medienpräsenz und Umfragewerte (siehe Abb.3), wobei die Umfragewerte immer höher liegen als die wordcounts. Die durchschnittliche Differenz der Trends beträgt -9,65 Prozentpunkte. Die Umfragewerte folgen also eindeutig nicht der medialen Präsenz.

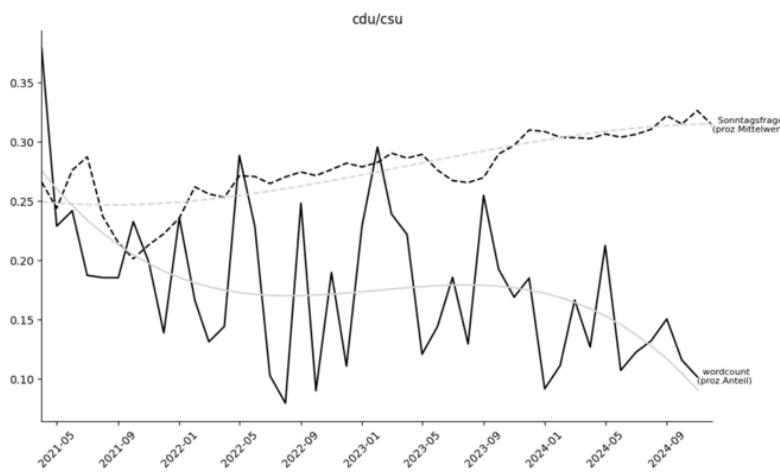


Abbildung 3: Anteil wordcounts vs. Umfragewerte CDU sowie Trends

Dagegen stimmen die Trends bei der SPD nahezu vollständig überein. Es ist allerdings nicht eindeutig, ob die Umfragewerte der Präsenz folgen, eher steigen/sinken beide nahezu

gleichzeitig und liegen immer nah beieinander (siehe Abb.4). Die durchschnittliche Differenz der Trends liegt bei 0,59 Prozentpunkten.

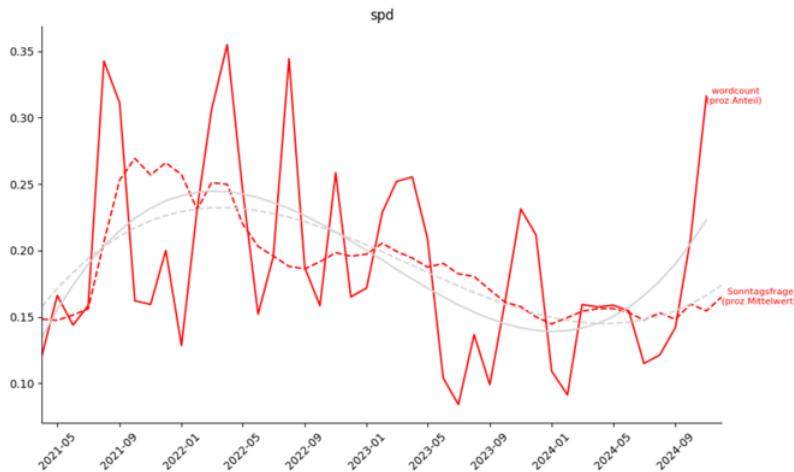


Abbildung 4: Anteil wordcounts vs. Umfragewerte SPD sowie Trends

Bei den Grünen ähneln sich die Trends ebenfalls von der Tendenz und nähern sich über den Verlauf einander an, grundsätzlich liegt der Trend der wordcounts für die meisten Monate höher als der der Umfragewerte, aber auch hier ist nicht eindeutig, wer wem folgt. Trotz zeitweise recht hohen wordcounts, sind nachfolgend nur Anfang 2022 steigende Umfragewerte zu beobachten, dann sinken die Umfragewerte wieder, trotz erneut immer wieder hoher wordcounts zwischen Sept. 2022 und Mai 2023. Die durchschnittliche Differenz der Trends liegt bei 6,55 Prozentpunkten.

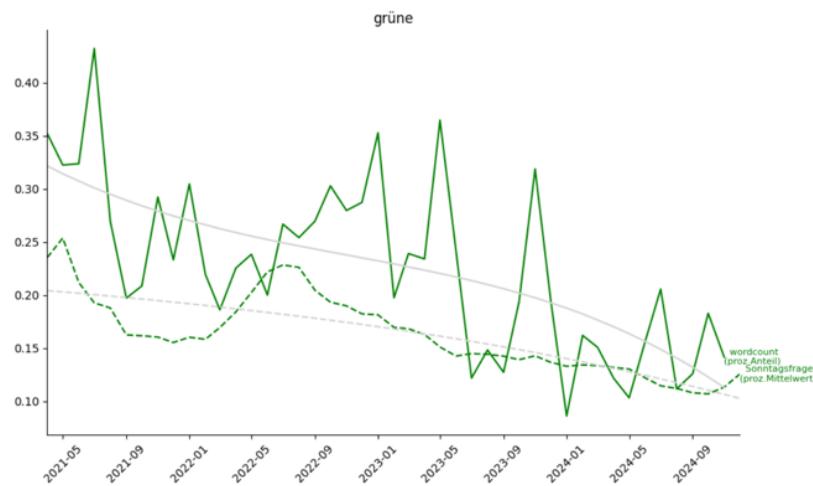


Abbildung 5: Anteil wordcounts vs. Umfragewerte Grüne sowie Trends

Die FDP erreicht besonders Ende 2021 und in 2022 schwankend immer wieder recht hohe wordcounts, sie liegen immer deutlich über den Umfragewerten. Nachfolgend ist nur einmal eine kleine Steigerung der Umfragewerte – nämlich im Oktober 2021 zu sehen, dann

nehmen aber die Umfragewerte immer weiter ab. Auch die wordcounts sinken ab Juni 2022 – es folgen hier also eher die wordcounts den Umfragewerten. Die durchschnittliche Differenz zwischen den Trends beträgt 9,18 Prozentpunkte.

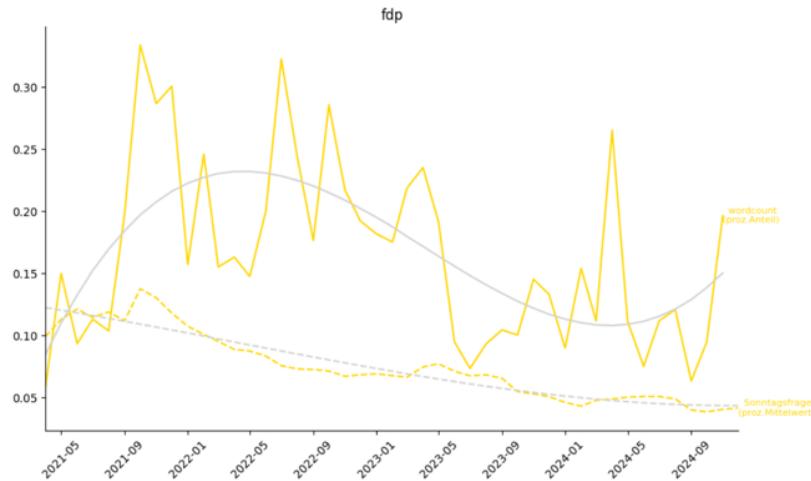


Abbildung 6: Anteil wordcounts vs. Umfragewerte FDP sowie Trends

Bei der AfD zeigen sich zunächst nah beieinander liegende Trends bis Ende 2022. Es sind immer wieder Spitzen bei der Medienpräsenz zu beobachten, während die Umfragewerte dezent steigen. Ab Mai 2023 steigen dann die wordcount-Spitzen von 20% auf Spitzen bis zu 60%. Eher parallel als nachfolgend werden diese von steigenden Umfragewerten begleitet. Die durchschnittliche Differenz der Trends beträgt für die AfD 7,96 Prozentpunkte. Die AfD ist die einzige Partei, die steigende Umfragewerte aufweist neben der CDU/CSU.

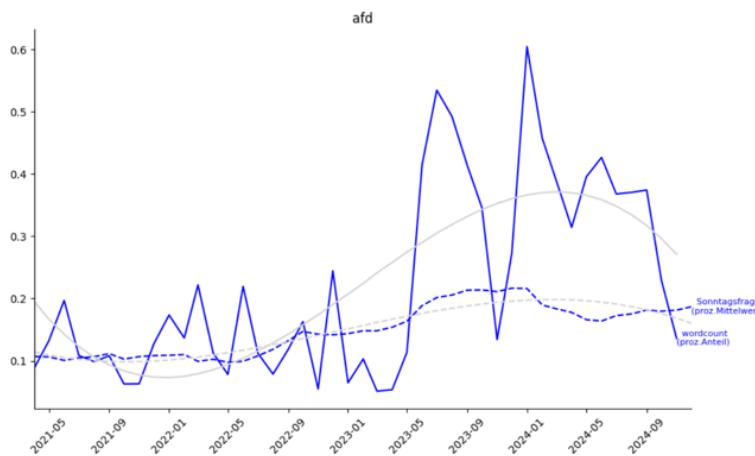


Abbildung 7: Anteil wordcounts vs. Umfragewerte SPD sowie Trends

## **Diskussion und Handlungsempfehlung**

Die Erwartung, dass auf eine steigende Medienpräsenz auch steigende Umfragewerte folgen, kann nicht eindeutig bestätigt werden. Dennoch lässt sich aber sagen, dass sich zumindest in der Tendenz (außer für die CDU/CSU) die Trends von Medienpräsenz und Umfragewerten gleichläufig verhalten und somit ein gegenseitiger Einfluss bestehen könnte. Mal steigt oder sinkt die Medienpräsenz zuerst, dann steigen oder sinken die Umfragewerte zuerst.

Gesondert zu betrachten ist die CDU/CSU. Als eine der beiden großen deutschen Parteien, ist die Bekanntheit sowieso als gesichert anzunehmen und sie scheint von der niedrigeren Medienpräsenz eher zu profitieren - vielleicht gilt für die Wähler das Motto: „Je weniger Schlagzeilen, desto verlässlicher.“ Dies könnte in einer qualitativen Untersuchung ggf. validiert werden.

Hervorzuheben ist ebenfalls die AfD als einzige weitere Partei mit immer steigenden Umfragewerten, sowie die Partei mit der höchsten medialen Präsenz ab Mai 2023. Die AfD ist eine ursprünglich kleinere Partei aus dem rechtspopulistischen Spektrum. Die Tendenz, dass solche Parteien mit großer medialer Präsenz auch bessere Wahlergebnisse erzielen, wurde in einer Untersuchung zu den spanischen Wahlergebnissen von 2019 beobachtet (siehe Ergebnisse von Estevez, Dominguez & Graña, 2019). In dieser Untersuchung wird auch von ähnlichen Beobachtungen aus Belgien und Dänemark berichtet. Diese Tendenz kann hier für Deutschland bestätigt werden. In einer weiterführenden Untersuchung könnten die Ursachen näher angeschaut werden, aber auch ob und weshalb dies nur für rechtspopulistische Parteien gilt und weshalb nicht Parteien wie die Grünen und die FDP von ihrer doch zeitweise ebenfalls hohen medialen Präsenz mehr profitieren.

Es könnte in einer weiterführenden Untersuchung zusätzlich ein Blick auf die mediale Präsenz von politischen Führungspersonen und ihren Einfluss auf die Umfragewerte und Wahlergebnisse geworfen werden. Außerdem könnte sich dann erneut mit geeigneten Methoden für den Vergleich und die Darstellung eines Zusammenhangs beschäftigt und diese eventuell angepasst oder erweitert werden.

## **Literaturverzeichnis / Quellen**

Estevez, J., Dominguez, JJ. & Graña, M. (2019). Relationship between the visibility of political leaders during campaign and the outcome in general elections. A case study for Spain. Abgerufen am 14.01.2024 von <https://arxiv.org/pdf/2002.07768.pdf> .

Github (2016). „german\_stopwords\_full“. Abgerufen am 15.01.2024 von [https://raw.githubusercontent.com/solariz/german\\_stopwords/master/german\\_stopwords\\_full.txt](https://raw.githubusercontent.com/solariz/german_stopwords/master/german_stopwords_full.txt)

Schulz, W. (2023). Analysen zur Wahlbeteiligung 2021: Medien und Wahlwerbung als Treiber der Wahlbeteiligung. Abgerufen am 14.01.2025 von [https://www.ard-media.de/fileadmin/user\\_upload/media-perspektiven/pdf/2023/MP\\_20\\_2023\\_Analysen\\_zur\\_Bundestagswahl\\_2021.pdf](https://www.ard-media.de/fileadmin/user_upload/media-perspektiven/pdf/2023/MP_20_2023_Analysen_zur_Bundestagswahl_2021.pdf) .

Wikipedia (2022). „Screen Scraping“. Abgerufen am 17.01.2024 von [https://de.wikipedia.org/wiki/Screen\\_Scraping](https://de.wikipedia.org/wiki/Screen_Scraping) .