## Projekttitel: [Wandel der Worte - Langzeitdatenanalyse journalistischer Perspektiven](https://wv.jugend-forscht.de/projects/edit/115225/)

Teilnehmende: Levi Blumenwitz (17)

Erarbeitungsort: Privat/zu Hause

Projektbetreuende: Melanie Mestl

Thema des Projekts: Analyse der Zeitungen „The New York Times“ und „The Guardian“

Fachgebiet: Arbeitswelt

Wettbewerbssparte: Jugend forscht

Bundesland: Bayern

Wettbewerbsjahr: 2025

*Ihr könnt euer eigenes Deckblatt erstellen und es gestalten. Das Deckblatt zählt nicht zu den 15 Seiten. Ein Deckblatt wird später auch noch einmal automatisch von Jugend forscht erstellt.*

**Gliederung der schriftlichen Arbeit**

|  |
| --- |
| Projektüberblick *(zählt nicht zu den max. 15 Seiten)* |
| Inhaltsverzeichnis *(zählt nicht zu den max. 15 Seiten)*   1. Fachliche Kurzfassung |
| 1. Motivation und Fragestellung 2. Hintergrund und theoretische Grundlagen |
| 1. Vorgehensweise, Materialien und Methoden |
| 1. Ergebnisse |
| 1. Ergebnisdiskussion |
| 1. Fazit und Ausblick |
| 1. Quellen- und Literaturverzeichnis *(zählt nicht zu den max. 15 Seiten)*   Unterstützungsleistungen |
|  |

**Projektüberblick**

Wie haben sich die Medien über die letzten Jahre verändert? In meinem Projekt befasse ich mich mit der Datenanalyse von zwei bedeutenden Zeitungen. Ausgewählt habe ich die "New York Times" (USA) und "The Guardian" (GB). Das Ziel ist, Artikel der beiden Zeitungen zu analysieren und auszuwerten, wobei ich mich bei den beiden Zeitungen auf jeweils drei Rubriken beschränkt habe. Die drei Themen - World, Politics und Opinion - sind bei beiden Zeitungen vergleichbar, und ich werde alle Artikel dieser Rubriken analysieren. Kriterien bei der Analyse sind aktuell die Objektivität und Polarisation der Artikel, sowie deren Länge und Anzahl. Der geplante Zeitraum dieser Analyse ist 10 bis 20 Jahre.

Das Ziel dieser Forschungsarbeit ist die Überprüfung von Vorurteilen bezüglich der Veränderung in der Berichterstattung sowie, je nach Ergebnis, die Glaubwürdigkeit von Qualitätsmedien zu stärken beziehungsweise die Veränderung der Medien im Laufe der Zeit nachzuweisen.

1. **Fachliche Kurzfassung**

Das Projekt untersucht die zeitliche Entwicklung von Sentiment (Polarisation und Objektivität), Artikellänge und Artikelanzahl in "The New York Times" und "The Guardian" zwischen 2010 und 2021. Mithilfe von Python wurden Artikel gesammelt und analysiert, um langfristige Trends zu identifizieren.  
Da für die Auswertung der Analyse eine Vielzahl von Optionen möglich ist, wurde zur Veranschaulichung exemplarisch die Analyse von Artikelanzahl…..?? folgende Ergebnisse ausgewählt.

Die Ergebnisse zeigen:

Die Artikelanzahl sinkt im Guardian in der Rubrik "Opinion", während sie im Politikbereich der New York Times deutlich ansteigt.  
Das Sentiment bleibt in beiden Zeitungen konstant (durchschnittlich neutral). Jedoch zeigt die Kategorie "World" eine höhere Subjektivität als "Politics" und "Opinion".  
Die durchschnittliche Artikellänge unterscheidet sich: NYT-Artikel sind mit 1100 Wörtern länger als Guardian-Artikel (800 Wörter), ohne signifikante Entwicklung über die Jahre.  
Diese Erkenntnisse verdeutlichen redaktionelle Unterschiede und zeigen, wie sich globale Ereignisse auf die Berichterstattung beider Zeitungen auswirken.

1. **Motivation und Fragestellung**

Die Diskussion über Fake News und das schwindende Vertrauen in die Medien ist in allgegewärtig. Laut **(**[**statita.com**](http://statita.com/)**)** Edeman Trust Barometer 2024 zeigt, fass Menschen eher in klassische Medienals in soziwle Medien vertrauen Jedoch hat „Deutschland, das Land der Qualitätspresse …auch ein Problem mit dem Vertrauen in die Leitmedien…nur 46 % gaben an, sie hätten Vertrauen in die Presse „  
…laut einer FORSA Umfrage aus dem Jahr 2022 sagen 43% der Journalismus sei schlechter geworden (Quelle) , “. Auch in den USA geben laut einer Studie aus dem Jahr 2023 gaben 60 % der Befragten an, ihr Vertrauen in die Medien sei gesunken.  
Laut (Quelle) können Leitmedien der Versuchung zu Polarisieren(…) noícht widerstehen und bringen (..)m die Demokratie in eine schwierige Lage…

Doch kann man auch wissenschaftlich untersuchen, ob sich die Berichterstattung tatsächlich ins Negative verändert hat?  
Als Nutzer von sozialen Medien war meine ursprüngliche Idee Artikel der Plattform X (ehemals Twitter) zu analysieren.

Nach ein wenig durchsuchen der Webseite viel mir schnell auf, dass die meisten Meinungen auf X in Form von Videos und Bildern dargestellt wurden. Da dies als Textanalyse nicht umsetzbar war, ging meine Suche weiter zu Facebook, wo das Auslesen der Artikel nicht unterstützt wurde, und zu Reddit, wo ein ähnliches Problem wie bei X auftrat. Daraufhin habe ich meinen Blick auf die sogenannten Leitmedien gerichtet/geworfen.

Bei meiner ersten Recherche stieß ich auf eine Studie von Michael Haller aus dem Jahr 2017 Mama?. Er hat die Berichterstattung von drei deutschen Zeitungen über einen Zeitraum von 20 Wochen untersucht. Er wertete 480 Zeitungsausgaben mit 2240 Seiten aus und analysierte 1687 Berichte und Kommentare.  
Meine Idee war eine Langzeitanalyse durchzuführen, um die Entwicklung der Medien über die Jahre hinweg zu analysieren. Mithilfe moderner Technologien wollte ich eine Analyse über einen viel größeren Zeitraum durchführen.

Hier habe ich zuerst mein Blick auf die "New York Times" geworfen. Diese ist bekannt für ihre objektive Berichterstattung und ist eine der größten Zeitungen in den USA.

Um aber die Werte objektiv zu betrachten und vergleichen brauchte ich eine zweite Zeitschrift, die meinen Anforderungen entsprach. Nach ein wenig suchen stoß ich auf "The Guardian", welche einen guten Kontrast bildet.  
Oder:  
Als zweite Zeitschrift haben ich the Guardian aus UK als Vergleich ausgewählt, ebenfalls wertgeschätzt für seinen unabhängigen Journalismus.

Ziel meiner Arbeit ist die Überprüfung folgender Fragestellungen:

Wie objektiv sind die beiden analysierten Zeitungen wirklich?

Wie stark hat sich die Medienberichterstattung bezogen auf Artikellänge und anzahl in den letzten Jahren verändert?

Wird die Medienberichte tatsächlich immer negativer dargestellt?

Sind Medien in den USA subjektiver als in Großbritannien?

1. **Hintergrund und theoretische Grundlagen**

Hier ist Platz für eine kurze Einordnung eures Themas in den wissenschaftlichen Zusammenhang.

* Was ist zu dem Thema schon bekannt?
* Wo knüpft eure Arbeit an?

**Fehlt noch**

1. **Vorgehensweise, Materialien und Methoden**

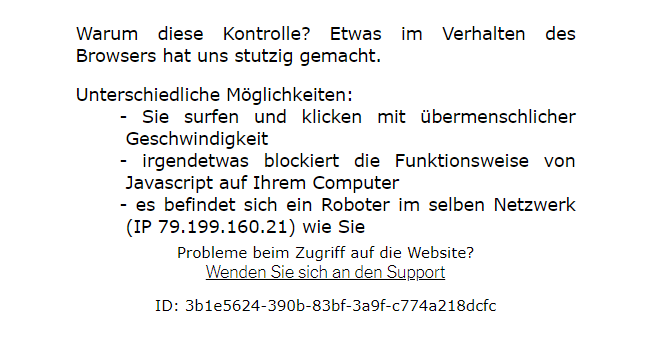
Der Vorgang, um die Daten zu sammeln und zu analysieren, ist sehr komplex und wird in mehreren Schritten durchgeführt.

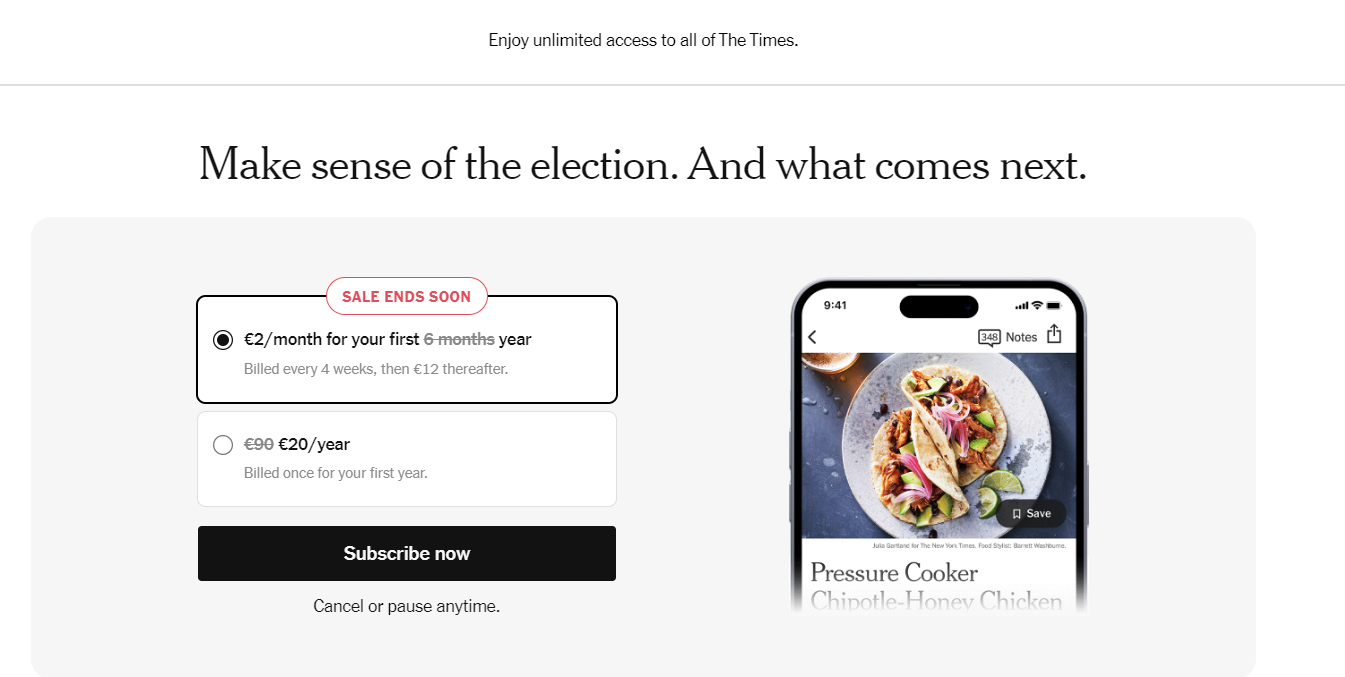
4.1. Links sammeln

Der erste Hauptschritt ist, den Artikeltext zu bekommen. Dafür muss ich als erstes Zugriff auf die kompletten Links der beiden Zeitschriften zugreifen. Dafür benutze ich die API von "The New York Times" und "The Guardian". Da nur sehr wenige Zeitschriften so eine API haben, musste ich mich auf die beiden Zeitschriften beschränken. Die API gibt mir die Möglichkeit mithilfe von der Bibliothek requests, die Links der Artikel zu bekommen und in einer Datei zu speichern.  
Aufgrund der Unterverzeichnisse des Links kann man das Datum, sowie Rubrik des Artikels auslesen und sortieren (Beispiel?).  
Hier habe ich mich für drei Rubriken entschieden, welche bei beiden Zeitschriften vergleichbar sind. Diese Rubriken sind "World", "Politics" und "Opinion". Jetzt sortiere ich die Links nach Datum und Rubrik.

Mithilfe der Links kann ich jetzt auf die Webseiten zugreifen. Doch um die Artikel zu analysieren, brauche ich Zugriff auf den Artikeltext. Dies wird unterteilt in zwei große Schritte:

4.2. Quellcode herunterladen

Als erstes benötige ich den Quellcode der Webseite. Das beschaffen des Quellcodes war der vermutlich aufwendigste Prozess der ganzen Arbeit. Der Quellcode ist der HTML-Code der Webseite, welcher alle Informationen der Webseite enthält. Diesen Code kann ich durch verschiedene Methoden herunterladen. Der Prozess ist sehr unterschiedlich, je nachdem welche Webseite ich herunterlade.  
Bei "The Guardian" war dieser Prozess viel einfacher. Ich konnte mit einer einfachen Anfrage mit dem Python Modul **requests** den Quellcode der Webseite herunterladen. Dieser wurde dann in einer Textdatei gespeichert, sortiert nach Datum und Rubrik.  
Bei der New York Times war es viel komplizierter. Es gibt verschiedene Methoden um den Quellcode herunterzuladen. Die Methode **requests** wie ich bei "The Guardian" genutzt habe, hat nicht funktioniert. Nach bereits 100 Artikeln wurde meine IP-Adresse blockiert (siehe Abbildung 1).

Die zweite Methode ist die Python Bibliothek **selenium**, welche eine beliebte Methode ist um einen echten Browser wie Chrome zu simulieren. Doch auch hier gab es Probleme. Die New York Times hat schnell meine ungewöhnliche Aktivität bemerkt, und nur den ersten Absatz des Artikels angezeigt. Außerdem wurde der Inhalt des Artikels hinter einer Paywall versteckt (siehe Abbildung 2).

Dies hat meine Analyse unmöglich gemacht, und ich musste einen Weg finden, um die Paywall, sowie die vielen Captchas zu umgehen.  
Mein erster Versuch war die Maßnahmen der New York zu verhindern, indem ich durch rotieren meiner Proxy die Captchas umgehe. Leider haben gratis Proxys oft nicht funktioniert. Eine andere Methode war das benutzen einer Externen API, zum Beispiel **Scraperapi**. Doch aufgrund limitierten Tokens in der Freemium Version, war dies auch keine Lösung. An diesem Punkt dachte ich das wäre das Ende meiner Arbeit, doch nach ein wenig herumprobieren, habe ich entdeckt, dass der komplette Artikeltext auch im "Backend" vorhanden ist. Aber leider in einem komplizierten Geflecht aus JSON-ähnlichen Strukturen. Das heißt ich konnte den Quellcode mit meiner vorher genannten Methode **Selenium** herunterladen und nachträglich den Text extrahieren.  
Da dieser Schritt am zeitaufwendigsten war, habe ich nach Wegen gesucht um diesen Prozess zu minimieren.

Meine erste Idee war das Projekt auf **Google Colab** als **Jupyter Notebook** auszuführen. Hier kann man nämlich Python Code online ausführen, sowie auch die **Selenium**-Methode nutzen. Aufgrund des weniger Zugriffmöglichem Dateisystem, habe ich dies jedoch verworfen. Meine zweite Idee war das Projekt auf einem **AWS (Amazon Web Services)** Server auszuführen. Hierbei habe ich einen AWS EC2-Server eingerichtet, um meine Anwendung zu hosten. Mehr dazu kann in meinem Medium Blog gefunden werden.

4.3. Text extrahieren

Hierfür habe ich eine Funktion erstellt, die den Text Stück für Stück den Text aus dem Backend herausfiltert und dann zusammengefügt. Mithilfe von **regex** konnte der Text gefiltert werden und die relevanten Daten gespeichert werden.Auch das extrahieren des Textes war einfacher bei "The Guardian". Hier habe ich einfach die herkömmliche Methode **BeautifulSoup** genutzt um anhand von HTML-Tags den Text zu lokalisieren. Dieser Text wurde dann in einer Textdatei gespeichert, sortiert nach Datum und Rubrik.

4.4. Text analysieren

Da ich nun den Artikeltext habe, kann ich diesen nach verschiedenen **??** analysieren.  
Zum Speichern der ausgewerteten Daten habe ich eine SQL-Datenbank erstellt. Hierfür benutze ich über das ganze Projekt hinweg Pythons Bibliothek SQLite3 benutzt. In dieser Datenbank speichere ich die Daten nach Datum und Rubrik, um später die Daten einfach abrufen zu können.

4.4.1. Wörteranzahl

In meinem Code wird dies als "Wordcount" bezeichnet, und es ist ziemlich selbstverständlich. Zuerst habe ich den Text in einzelne Wörter aufgeteilt und diese gezählt. Diese Daten habe ich dann in einer SQL-Datenbank gespeichert.

4.4.2. Sentimentalanalyse

Die Sentimentanalyse ist ein wichtiger Bestandteil meiner Arbeit. Hierbei wird der Text auf Polarisation sowie Subjektivität hin analysiert. Dies wird mithilfe des Moduls**TextBlob** durchgeführt.

**4.4.2.1. Polarisation**

Dieses Modul gibt jedem Wort eine Wertung von -1 bis 1, wobei -1 negativ und 1 positiv ist. Daran kann man erkennen, ob ein Text positiv oder negativ ist.

**4.4.2.1. Subjektivität**

**TextBlob** berechnet die Subjektivität, indem es die 'Intensität' betrachtet. Die Intensität bestimmt, ob ein Wort das nächste Wort modifiziert.

Diese beiden Werte werden anhand des Datums als Indikator in einer SQL-Datei gespeichert, um späteres Abrufen zu erleichtern.

4.4.3. Artikelanzahl

Hierbei wird die Anzahl der Artikel von jedem Tag gezählt und in einer SQL-Datei gespeichert.  
Dadurch kann man die Entwicklung der Artikelanzahl über die Jahre hinwegsehen.  
Hier wird unterteilt in die Anzahl der Artikel pro Tag sowie pro Monat.

4.5. Graphen erstellen

Die Daten, die ich in den SQL-Dateien gespeichert habe, werden in einem Graphen dargestellt.  
Hierfür habe ich als erstes eine globale Funktion erstellt, die mithilfe von verschiedenen Eingabeparametern den Graphen erstellt. ([Plotting\Plotting.py](https://file+.vscode-resource.vscode-cdn.net/c:/Users/L-Blu/Levi/Programmieren/Python/Jugend-Forscht/Plotting/Plotting.py)).  
Diese Funktion wird dann in den einzelnen Dateien aufgerufen, benötigt zum Zeichnen des Graphen werden die Daten aus der Datenbank, alle Spalten, die entfernt werden sollen (z.B. ID), und andere Kleinigkeiten wie der Titel und Farben benötigt.  
Innerhalb der verschiedenen Analysetypen erstelle ich weitere Wiederholungsanweisungen um die für jede Rubrik einen Graphen zu erstellen, sowie für jedes Jahr, um die Entwicklung über die Jahre und innerhalb der Jahre zu analysieren.

4.6. Interaktive Webseite erstellen

4.6.1. Datenauswahl

Ein weiterer Weg um meine Daten zu visualisieren war durch erstellen einer interaktiven Webseite. Diese Webseite wurde mit Streamlit erstellt. Streamlit ist ein sehr beliebtes Modul um Webseiten in Python zu erstellen. Auf der Webseite kann man als erstes auswählen, welche Option man analysieren möchte (Wordcount, Polarisation, Subjektivität, Artikelanzahl). Danach kann man die auswählen, wie viele verschiedene Graphen man übereinander angezeigt haben möchte. Je nach Anzahl kann man auswählen was für Daten der jeweilige Graph haben soll. Nach auswahl der Daten kann man auswählen in welchem Jahr-Zeitraum? die Daten angezeigt werden sollen.

4.6.2. Graph erstellen

Jetzt kann man den Graphen erst durch drücken eines Buttons erstellen, um andauerndes Laden des Graphen im Hintergrund zu vermeiden. Der Graph wird mithilfe der Interaktiven Bibliothek plotly erstellt, die es ermöglicht, den Graphen zu zoomen, zu verschieben und bestimmte Daten mithilfe der Legende auszublenden.

4.6.3. Tabelle mit Top 10

Zusätzlich zu den Graphen habe ich eine Tabelle erstellt, welche den Datensatz nach den in [6.2.](#61-datenauswahl) ausgewählten Kriterien sortiert und die Top 10 anzeigt.  
Angezeigt wird dann der Titel des Artikels, das Datum und auch die Content ID des Artikels. Mit dieser ID kann man auf einer anderen Webseite den Artikel direkt aufrufen, und genauer anschauen. Diese Webseite ist öffentlich nicht zugänglich, da die Daten nur Lokal auf meinem Laptop gespeichert sind.

1. **Ergebnisse** (etwa vier Seiten)

Ziel der Arbeit war eine Langzeitdatenanalyse aller Artikel der beiden Zeitschriften. Aufgrund der zeitlichen Begrenzung in dieser Forschungsarbeit konnte ich nicht die kompletten Daten herunterladen. Deshalb habe ich mich darauf beschränkt die Entwicklung von 2010 - 2021 genauer zu analysieren, indem ich den Anfang des Zeitraums 2010 - 2011 und das Ende 2020 - 2021 heruntergeladen habe. Mit diesen Daten kann ich nun abschätzen, wie sich die verschiedenen Aspekte entwickelt haben.

Ich habe zum Vorstellen der Ergebnisse besonders aussagekräftige Ergebnisse ausgewählt, die die Entwicklung der beiden Zeitungen gut darstellen. Man kann alle anderen Graphen in der interaktiven Webseite nachschauen, sowie auf den generierten Graphen. (Fußnote?)

**1. Entwicklung der Artikelanzahl**

Auf dem dargestellten Bild kann man die Entwicklung der Artikelanzahl der Rubrik "Opinion" der Zeitung "The Guardian" sehen. Jeder Punkt stellt einen Tag dar. Die Y-Achse zeigt die Anzahl der Artikel, die an diesem Tag veröffentlicht wurden. Man erkennt, dass in der Mitte des Graphen keine Punkte vorhanden sind. Dies ist aufgrund der fehlenden Daten in diesem Zeitraum. In schwarz ist die Regressionsgerade dargestellt, welche die Entwicklung der Artikelanzahl über die Jahre hinweg darstellt. Wie auch an dem Text in der recht oberen Ecke zu erkennen ist, ist die Entwicklung der Artikelanzahl in den letzten Jahren abgesunken. Durchschnittlich sind es ca. 11 Artikel weniger im Jahr 2021 als im Jahr 2010.

[Artikelanzahl - Guardian - Opinion - Days]



Hingegen ist bei der New York Times ein starker Anstieg zu erkennen im Bereich Politik. Mit einem Anstieg von ca. 8 Artikeln ist dies auch visuell im Graphen wiederzuerkennen.

[Artikelanzahl - NYT - Politics - Days]



Bestätigt wird dies durch die Graphen der Artikelanzahl pro Monat. Hierbei ist ein starker Anstieg bei der New York Times zu erkennen, während bei "The Guardian" ein starker Abfall stattfindet.

[Artikelanzahl - Guardian - Opinion - Month]



[Artikelanzahl - NYT - Politics - Days]

**2. Entwicklung des Sentiments**

Die Sentimentanalyse wird unterteilt in Polarisation und Subjektivität.

**Die Polarisation** gibt an, ob ein Text positiv oder negativ ist. -1 bedeutet sehr negativ, 1 bedeutet positiv. Ein durschnittlicher Wert beider Zeitschriften ist 0.1.

[Polarisation - Guardian - Together]

  
In der Grafik werden die 3 Kategorien der Zeitschrift "The Guardian" verglichen. Es ist klar dargestellt, dass die Polarisation über die Jahre hinweg relativ konstant ist und auch zwischen den Kategorien vergleichbar ist. Es gibt keine großen Schwankungen, und die Werte sind immer um ca. 0.1.

[Polarisation - NYT - Together]

  
Auch bei der New York Times ist ein ähnliches Bild zu erkennen. Die Werte sind relativ konstant und auch hier gibt es keine großen Schwankungen. Die Werte sind immer um ca. 0.1.

**Die Subjektivität** gibt an, wie subjektiv ein Text ist. 0 bedeutet sehr objektiv, 1 bedeutet subjektiv. Ein durschnittlicher Wert beider Zeitschriften ist 0.4, wie man an der Grafik unten erkennen kann. Hier wurden alle drei Kategorien zusammengefasst und in einem Graphen dargestellt.

[Subjektivität - NYT - All]



Wenn man aber die einzelnen Kategorien betrachtet (Bild unten), sieht man deutliche Unterschiede. Die Kategorie "World" ist deutlich subjektiver als die anderen beiden Kategorien. Dies ist auch bei "The Guardian" zu erkennen. Die Kategorie "World" ist zu beginn deutlich subjektiver als die anderen beiden Kategorien, und die Kateogrie "Opinion" ist am subjektivsten.

[Subjektivität - NYT - Together]



Interessant ist auch, dass dieses Ergebnis sehr vergleichbar mit dem Ergebnis von "The Guardian" ist. Auch hier ist die Kategorie "World" am subjektivsten, wobei sich "Politics" langsam vom eher Objektiven "Opinion" entfernt und sich "World" nähert:

[Subjektivität - Guardian - Together]



**3. Wörteranzahl & Artikellänge**

Zuletzt noch die Wörteranzahl und die Artikellänge. "The New York Times" hat eine deutlich höhere durschnittliche Wörteranzahl mit ca. 1100 Wörtern pro Artikel. Hingegen hat "The Guardian" mit einer durschnittlichen Wörteranzahl von 800 Wörtern.  
Abgesehen von dem durschnittlichen Unterschied der Wörteranzahl, gibt es auch Entwicklungen über die Jahre. Doch diese sind nicht sehr aussagekräftig, da die Korrelation sehr gering ist, was am Korreleationskoeffizienten von nur 0.09 zu erkennen ist. (Bild unten)

[Wordcount - Guardian - All]



1. **Ergebnisdiskussion** (etwa zwei bis drei Seiten)

Polarisation von 0,1 ist ein positives Signal, es zeigt, dass es in den untersuchten Medien keine starken Spaltungen gibt, sowie das gesellschaftliche und politische Diskussionen gemäßigt bleiben.  
Der Anstieg der Artikelanzahl ind der Rubrik Politik bei "The New York Times" und der Rückgang von was nochmal lassen Möglichkeiten in der Interpretation. Mögliche Ursachen könnten sein, dass die NYT aufgrung großer politischer Themen (trump/Biden/SkandaleWahlen…Ereignisse…) mehr den Fokus auf politische Berichterstattung gelegt hat, während sich "The Guardian" zum Beispiel mehr auf andere Themen wie etwa dem Klimawandel etc fokussiert haben könnte…  
Um dies weiter genauer zu analysieren könnte man zum Beispiel die Daten spezieller Jahre (Wahljahre/Sturm des Capitol/Brexit) herausfiltern..  
Man kann sehen, dass sich die Medien an unterschiedliche Anforderungen…anpassen.

1. **Fazit und Ausblick** (max. eine Seite)

Das Ziel wurde erreicht, jedoch konnte aufgrund der begrenzten Zeit nur ein Teil der Daten heruntergeladen werden.  
Es konnte eine Veränderung der Medien über den Zeitraum von 10 Jahren festgestellt werden, jedoch konnte die Vermutung der zunehmenden Polarisierung, also extremeren Meinung hier noch dran arbeiten, nicht bestätigt werden.  
Eine konstante Polarisierung hat verschiedene Vorteile, denn durch eine konstante Polarisation kann das Vertrauen der Leser der Zeitungen gestärkt werden.  
Auch beim Vergleich beider Zeitungen wird deutlich, dass die Berichterstattung in beiden Zeitungen ähnlich ist. Dies zeigt auf eine seriöse Berichterstattung beider Zeitungen, unabhängig vom Herkunftsland.  
Ein deutlicher Unterschied zwischen der amerikanischen Zeitung "The New York Times" und der britischen Zeitung "The Guardian" konnte in dieser Forschungsarbeit nicht festgestellt werden.  
Mein Projekt liefert ein konkretes Beispiel, wie Datenanalyse genutzt werden kann, um Vorurteile und Behauptungen über Medien zu überprüfen. Mit meiner Forschungsarbeit wurde für zukünftige Untersuchungen ein Tool erstellt, welches genutzt werden kann, um Artikel der beiden Zeitungen herunterzuladen und zu analysieren.  
Diese Analyse kann auf weitere Zeitungen ausgeweitet werden, um so eine umfassende Analyse der Medienlandschaft zu ermöglichen.  
Außerdem kann die Analyse auf weiter Jahre erweitert werden um Trends und Entwicklungen über einen längeren Zeitraum zu untersuchen.  
Auch die verschiedenen Analysetools können erweitert werden. Weitere Ideen waren das Analysieren der Artikel auf Schlagwörter, um so die Themen der Artikel zu identifizieren, und dadurch die Entwicklung und Relevanz von Themen über die Jahre hinweg zu analysieren. Eine weitere Option war die Untersuchung der Komplexität der Artikel Zitat, dass Menschen dummer werden? um so die Verständlichkeit der Artikel zu analysieren.  
Mein Projekt bietet eine Vielzahl an Möglichkeiten für Unternehmen, Trendanalyse, Wettbewerbsvorteile Bildungeinrichtungen Medienkritik, überprüfung von Vorurteilen, Auswahl verlässlicher Medien um gezielte Recherchen vorzunehmen.  
Durch die Webseite kann die Auswahl der Daten speziell gefiltert werden um so spezielle Fragestellungen zu untersuchen.  
Je nach Wunsch können so spezielle Fragestellungen untersucht und analysiert werden.  
Mit meiner Forschungsarbeit möchte ich einen Beitrag zur Medienkritik leisten, denn "in einer von Medien geprägten Welt [...] kann es gar nicht genug Medienkritik geben"Zitat und die "Mediendebatte ist in letzter Instanz auch [...] eine Debatte über den Zustand und die Zukunft der Demokratie"Zitat.

1. **Quellen- und Literaturverzeichnis** (zählt nicht zu den max. 15 Seiten)

**1. Python als Programmiersprache**

<https://www.python.org/:> 31.12.24, Python Software Foundation, Python als Programmiersprache

**2. Verwendete Python Bibliotheken**

<https://www.crummy.com/software/BeautifulSoup/:> 30.12.2024, Leonard Richardson, Beautiful Soup for HTML-Parsing

<https://matplotlib.org/:> 30.12.2024, © 2012 – 2024 The Matplotlib development team, Matplotlib für Graphen

<https://numpy.org/:> 30.12.2024, © 2024 NumPy team, Numpy für numerische Berechnungen

<https://pandas.pydata.org/:> 30.12.2024, © 2024 pandas, pandas zur Datenanalyse

<https://plotly.com/>: 30.12.2024, © 2024 Plotly, Plotly für interaktive Graphen

[https://requests.readthedocs.io/en/master/:](https://requests.readthedocs.io/en/master/) 30.12.2024, © 2024. A Kenneth Reitz Project, Requests: HTTP for Humans

<https://scikit-learn.org/>: 30.12.2024, © 2007 - 2024 scikit-learn developers (BSD License), scikit-learn for Regression

<https://www.selenium.dev/>: 30.12.2024, © 2024 Selenium Software Freedom Conservancy, Selenium für Web-Scraping

<https://docs.python.org/3/library/sqlite3.html>: 30.12.2024, SQLite3 für Datenbanken

[[https://docs.python.org/3/library/re.html:](https://docs.python.org/3/library/re.html)](https://www.sqlite.org/:) 30.12.2024, Python Software Foundation, Regular expression operations

[https://streamlit.io/:](https://streamlit.io/) 30.12.2024, © 2024 Snowflake Inc., Streamlit für Webseiten

[https://textblob.readthedocs.io/en/dev/:](https://textblob.readthedocs.io/en/dev/) 30.12.2024, © Steven Loria, TextBlob für NLP

**3. Webseiten**

[https://colab.google/:](https://colab.google/) 03.01.25, Google, Google Colab

[https://jupyter.org/:](https://jupyter.org/) 03.01.25, Jupyter, Jupyter Notebook

[https://www.theguardian.com/:](https://www.theguardian.com/) 03.01.25, © 2025 Guardian News, The Guardian

[https://www.nytimes.com/:](https://www.nytimes.com/) 03.01.25, © 2025 The New York Times Company, The New York Times

[https://aws.amazon.com/de/:](https://aws.amazon.com/de/) 03.01.25, 2024 Amazon Web Services Inc, Amazon Web Services

**4. Literatur**

**Die vierte Gewalt** – Wie Mehrheitsmeinung gemacht wird, auch wenn sie keine ist; 2022-09-28, Richard David Precht

**Lückenpresse** - Das Ende des Journalismus, wie wir ihn kannten; 2016-09-01, Ulrich Teusch

**Die große Gereiztheit** - Wege aus der kollektiven Erregung; 2018-02-19, Bernhard Pörksen

**Fake Facts** - Wie Verschwörungstheorien unser Denken bestimmen; 2020-05-15, Katharina Nocun

**Quellenangabe bei Internetseiten**

* Genaue URL (Webadresse) und Datum des Seitenaufrufs
* verantwortliche Person der Seite, Titel und Thema des Inhalts

*Beispiel: Stiftung Jugend forscht e. V., schriftliche Arbeit und weiterführende Informationen,* [*http://jugend-forscht.de/teilnahme/ablauf/schriftliche-arbeit.html*](http://jugend-forscht.de/teilnahme/ablauf/schriftliche-arbeit.html)*, besucht am 12.07.2022.*

**Quellenangabe bei Büchern**

* Vorname und Nachname der Autorin/des Autors/der Autoren
* Buchtitel
* Erscheinungsort
* Erscheinungsjahr
* Seitenangabe

*Beispiel: Andrea Gruß, Ute Hänsler: „Knallraketen und Gummigeister“, Frankfurt/Main 2020, S.10*

**Quellenangabe bei Zeitschriften**

* Zusätzlich zu den Angaben wie bei Büchern werden der Name der Zeitschrift, die Nummer der Ausgabe und die Seitenangabe des Artikels angegeben.

*Beispiel: Susanne Strunk: „Prozesse und Wirkungen der Teilnahme an Schulwettbewerben“ in Die Deutsche Schule, Zeitschrift für Erziehungswissenschaft, 104. Jahrgang, Heft 2, 2020, S. 150*

**Quellenangabe bei Fotos**

Quellenangaben für Fotos werden im Allgemeinen direkt unter das Foto gesetzt:

* Agentur, Unternehmen oder Institution
* Vorname und Nachname der Fotografin/des Fotografen

*Beispiel: Foto: Agentur Krummen, Robert Schnappschuss*

**Unterstützungsleistungen (Angabe in JufoWV!)**

Die vollständige Angabe von Unterstützungsleistungen ist wichtiger Bestandteil der Dokumentation eines wissenschaftlichen Projektes. Eine detaillierte Erläuterung von Unterstützungsleistungen hilft den Jurymitgliedern bei der Einschätzung des Eigenanteils eurer Arbeit und erleichtert das Jurygespräch. Vielfalt und Umfang von Unterstützungsleistungen variieren stark und bestimmen nicht die Qualität des Projektes. Eure Unterstützungsleistungen tragt ihr mit Abgabe eurer schriftlichen Arbeit **in JufoWV** ein. Dort gibt es ein eigenes Dropdown-Menü. Plant genügend Zeit für das Angeben eurer Unterstützungsleistungen ein.